AD-A214 143

SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

1a. REPORT SECURITY CLASSIFICATION UNCLASSIFIED	1b. RESTRICTIVE MARKINGS
2a. SECURITY CLASSIFICATION AUTHORITY	3. DISTRIBUTION/AVAILABILITY OF REPORT DISTRIBUTION UNLIMITED
2b. DECLASSIFICATION / DOWNGRADING SCHEDULE	
4. PERFORMING ORGANIZATION REPORT NUMBER(S)	5. MONITORING ORGANIZATION REPORT NUMBER(S)
6a. NAME OF PERFORMING ORGANIZATION Deputy Under Secretary of the Army (Operations Research) — 6b. OFFICE SYMBOL (If applicable) SAUS-OR	7a. NAME OF MONITORING ORGANIZATION
6c. ADDRESS (City, State, and ZIP Code) HQDA ATTN: SAUS-OR, RM 2E660 Washington, D.C. 20310-0102	7b. ADDRESS (City, State, and ZIP Code)
8a. NAME OF FUNDING/SPONSORING ORGANIZATION — 8b. OFFICE SYMBOL (If applicable)	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER
8c. ADDRESS (City, State, and ZIP Code)	10. SOURCE OF FUNDING NUMBERS PROGRAM ELEMENT NO. / PROJECT NO. / TASK NO. / WORK UNIT ACCESSION NO.

11. TITLE (Include Security Classification)

Pershing II Folow-On Test: Size Reduced by Sequential Analysis

12. PERSONAL AUTHOR(S)
Willard, Daniel

13a. TYPE OF REPORT	13b. TIME COVERED	14. DATE OF REPORT (Year, Month, Day)	15. PAGE COUNT
Final	FROM ______ TO ______	30 September 1984	

16. SUPPLEMENTARY NOTATION

17. COSATI CODES			18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
FIELD	GROUP	SUB-GROUP	Reliability; Sequential Analysis; Truncated Test; Bayesian Statistics; Fisher's Exact Test.

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

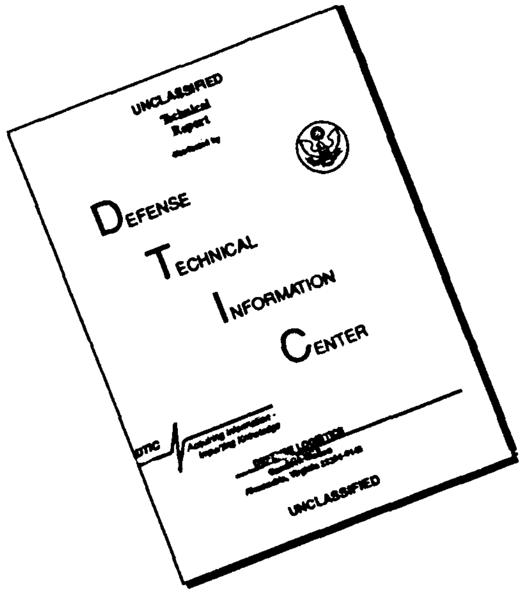
To replace the erroneous use of Fisher's Exact Test to derive the reliability of a missile system in annual tests, the author developed, in collaboration with statisticians, two efficient methods, both treated by frequentist and Bayesian methods. Particular attention is called to the Truncated Sequential Probability Ratio Test. Both methods gave essentially the same results, and so justified reduction in the size of the tests (for a given level of confidence) and therewith a reduction in the costs of the test program. The advantages of the Bayesian methods are pointed out. They are generally applicable for lot sampling and other uses of sequential tests.

20. DISTRIBUTION/AVAILABILITY OF ABSTRACT ☐ UNCLASSIFIED/UNLIMITED ☒ SAME AS RPT. ☐ DTIC USERS	21. ABSTRACT SECURITY CLASSIFICATION UNCLASSIFIED
22a. NAME OF RESPONSIBLE INDIVIDUAL Dr. Daniel Willard	22b. TELEPHONE (Include Area Code) (202) 697-6742 — 22c. OFFICE SYMBOL SAUS-OR

DD FORM 1473, 84 MAR — 83 APR edition may be used until exhausted. All other editions are obsolete.

SECURITY CLASSIFICATION OF THIS PAGE

DISCLAIMER NOTICE



THIS DOCUMENT IS BEST QUALITY AVAILABLE. THE COPY FURNISHED TO DTIC CONTAINED A SIGNIFICANT NUMBER OF PAGES WHICH DO NOT REPRODUCE LEGIBLY.

PERSHING II FOLLOW-ON TEST:

SIZE REDUCED BY SEQUENTIAL ANALYSIS

by

Daniel Willard

September 1984

SUBJECT: Pershing II Follow-On Test: Size Reduced by Sequential Analysis

By memorandum of 30 August 1982 (Reference 1), the Under Secretary of the Army tasked the service to "review our [operational test] methodology, to include considerations of mathematical rigor, risks, planning horizon, costs, and operational matters." In discussion of this matter with the author, he further elaborated the objectives:

a) Minimize cost of testing over the program life. Monitor all test results, including those of components as well as of the system, to minimize "no-tests" and to save on full-up tests. Use sequential analysis to further pare requirements for missile flights.

b) Criteria of test adequacy should be no more severe than those of other services (e.g., Minuteman, Poseidon).

c) Challenge the necessity for an annual update.

d) Consider whether testing, maintenance float, and reload were independent requirements as opposed to multiple missions for the same inventory of missiles.

The task was passed to the Army Research Office (Research Triangle, NC) which manages the business of the Army's Mathematics Steering Committee (Dr. Jagdish Chandra, Chairman), supporting mathematical research of relevance to the Army and the improvements in mathematical methods employed in the Army's research and study agencies.

The work summarized here is composed of contributions of several statisticians whose aid was solicited by the AMSC: Dr. Michael Woodroofe (University of Michigan)*, Dr. Nozer Singpurwalla (George Washington University), and Dr. Robert Launer (Army Research Office), as well as the author of this report. Others have provided informal comments and criticisms. An early version of this paper, prior to the author's knowledge of this other research, was presented as a talk at a conference of Army mathematicians (Reference 2).

* At Rutgers University during the course of this research.

Chapter I

The Problem

Two documents combined set forth the guidance the Joint Chiefs of Staff have provided to the military services regarding the conduct and reporting of tests of certain systems. For the Army only the Pershing Missile system is covered (Pershing I and Ia, and now Pershing II).

In a memorandum of 1975 (Reference 3), the Joint Chief of Staff directed that numerical confidence statements should be based on WSEG Report 92C (Reference 4), an extract of which is at Appendix C. "The goal of a test program should be to allow detection of a minimum change of X percent at the Y percent confidence level." * It suggests, by way of example, the use of Fisher's Exact Test to demonstrate success or failure in meeting this criterion.

References 3 and 4 have just been superseded. The revisions (References 5 and 6) eliminate an ambiguity and add considerations not previously called for and not discussed here except to note that the criteria to be applied to Pershing II are now less demanding than those applied to strategic systems. Fisher's Exact Test is still countenanced.

This use of this criterion appeared to the author to lack a sound statistical justification, and attempts to patch it up were unsuccessful. Appeal to a number of practicing statisticians within and outside the Army supported my challenge to Fisher's Exact Test (FET) in its application to Pershing reliability tracking. No one was contesting the ability of the FET to provide estimates of the probability that two samples, which have yielded pass-fail data, come from the same parent population, though Kendall and Stuart (Reference 7), do condemn its use for small samples.

With such an error apparently arising from an application of the methods of the "frequency" school of statistics, the obvious alternative was to try the methods of the "Bayesian" school.

There are many expositions of methods based on the use of Bayes' Theorem, the most recent of which--"Bayesian Reliability Analysis" by Martz and Waller--(Reference 8) I shall quote at intervals. Among the works arguing for the adoption of Bayesian methods, the following are noteworthy:

* X and Y are classified numbers.

Raiffa and Schlaifer - Applied Statistical Decision Theory (Reference 9) with a very complete description of the method of conjugate prior distributions.

Jaynes E.T., "Prior Probabilities" (IEEE Transactions on System Science and Cybernetics, September 1968) (Reference 10). Deduction from the principles of maximum entropy and invariance under certain group transformations leads directly to the Beta distribution as conjugate prior to a Bernoulli process; indeed to

$$dP(p;n,s) = p^{s-1}(1-p)^{n-s-1}\,dp \,/\, B(s,n-s) \qquad 1.1$$

where s is the number of successes in n trials observed as the basis for estimating p. This removes some of the "ad hoc" or "mathematically convenient" color of conjugate priors when relying on Raiffa and Schlaifer.

Martz and Waller perhaps epitomize the case best:

"There are several benefits in using Bayesian methods in reliability. First of all, it is important to recognize that all statistical inferential theories, whether sampling theory, Bayesian, likelihood, or otherwise, require some degree of subjectivity in their use. Sampling theory requires assumptions concerning such things as a sampling model, confidence coefficient, which estimator to use, and so on. For example, a sampling theory analysis proceeds as if it were believed a priori that the data were exactly [exponentially] distributed, that each observation had exactly the same mean life θ, and that each observation was distributed exactly independently of every other sample observation. The Bayesian method provides a satisfactory way of explicitly introducing and organizing assumptions regarding prior knowledge or ignorance. These assumptions lead via Bayes' theorem to posterior inferences, that is, inference obtained once the data have been incorporated into the analysis, about the reliability parameter(s) of interest. Bayes' theorem provides a simple, error-free truism for incorporating the prior information. The engineering judgment and prior knowledge are brought out into the open and are there for everyone to see instead of being quietly hidden. The engineer usually appreciates this opportunity to divulge such prior information in a formalized way."

The authors I commend are not, on philosophical matters, in complete agreement, and the authors (and critics) of the methods proposed in this paper have their differences, some of which become important as we proceed.

Suffice it to say that the Bayesian approach requires a more careful statement of the problem, to include in particular the prior distribution function, costs and risks: matters which the frequentists collapse into the confidence limits α and β. If there is indeed a legitimate uncertainty in (the form of) the prior distribution, that uncertainty must surely propagate into an uncertainty in the predictions for the process. In some cases results can be shown to be insensitive to the prior, and thus a convergence of Bayesian and frequentist answers occurs; but lacking such invariance, the frequentists are hard pressed to prove they have solved the right problem.

Having said this, I must confess that for some purposes we shall employ the frequentist approach, primarily because a full Bayesian solution has not been worked out.

Section 1. Literal Interpretation of JCS Guidance:

". . . annual . . . detection of a minimum reliability change of X percent at the Y percent confidence level."

A "change" in something means that its previous value has been defined. It would appear that an evaluation of the results of the first year's Follow-on-Test (FOT) is to be compared to that of the Operational Test (the base-line)(OT), and the evaluations of subsequent FOTs are to be compared to the evaluations made a year ago. The tests being of something less than the full combat mode of the system, projection to combat capability is to be made; thus while test results are to be reported, they are to be interpreted as well. This interpretation is surely to be based on all prior knowledge of system performance; i.e., all prior testing as well as that most recently at hand, "weighted" (one might say) by expert judgment of the relevance of older tests and analysis.

In the case of Pershing II, we shall have an inventory of missiles produced over a period of time and expected to be in service for a longer period. From the point of view of homogeneity, the inventory may need to be divided into two or more blocks, based on the significance of any changes in the production process during the run. When they are subjected to (annual) test, missiles will be of different ages as well from different blocks; so serial number and age may influence reliability at the time of testing or use in combat. It is clear, then, that in treating of a "change" in reliability, we are dealing with an uncertain base. Options which are open to us include:

a) Computing a "best" estimate from the OT firings, and treating it as the exact value of the reliability at that time of all the inventory.

b) Computing as in (a), but associating an uncertainty (standard deviation) to it also, to describe the uncertain reference point.

In either case, the results of each subsequent (annual) test would be compared to this as standard.

c) Computing as in (b), but then modifying the estimates using the results of subsequent tests (more trials, more successes, more failures). There are extremes in this process which are to be avoided:

(i) This modification might consist of using only the previous year's results as indication of the remaining inventory.

(ii) This modification might consist of accumulating the results of all prior tests, without regard to the aging effect or block modifications.

Judgment is clearly needed. Limiting the criterion to the smallness of the latest annual change (with small samples in the two cases) could result in a dangerous accumulation of change over the system life. On the other hand, where no statistically significant change has been detected, it would be reasonable to add one year's results to the results of the whole prior test series of a homogeneous block in estimating the average value at, say, the average age of the tested articles. It is probably not possible to specify in advance the details of the critical results to be reported. What is more important is that analyses be conducted to discover what are the constant and what are the variable components of the system reliability. Finally, detection of a trend should make it possible to forecast when the results of that trend will no longer be tolerable, and so signal the degree of urgency with which management should act to correct the trend.

d) This brings us to the question of the frequency of reporting the results of testing and analysis. The current practice is an annual report which probably has its roots in adminstrative cycles. The technical problems which reporting communicates to management are probably of two sorts: long-term aging with gradual deterioration, ("one-hoss shay" syndrome) and catastrophic failures. The latter tend to announce their presence in consistent repetitions of particular failure modes, and so call for out-of-cycle action no matter what the standard interval between reports. The former, on the other hand, are evidence of problems only slowly exacerbating, and so allow a more leisurely pace of administrative response. Alternatives to the present annual cycle are proposed below, for situations in which no guarantee of a clear bright green light or red light is available annually: (i) A guarantee can be given of a low likelihood of having to wait more than, say, 16 months for such a signal, along with the provision of a technical review of all failures showing any repetitions of mode. (ii) Administratively, skipping one year's report may be simpler.

These options will be explored in one or more places in the mathematical sections to follow.

Two assumptions have immediately to be disposed of:

1) Because Fisher's Exact Test is mentioned in JCS guidance, its use is correct and mandatory.

Fisher's Exact Test is an enumeration of all possible relative outcomes in two series of pass-fail tests, subject to the restraints that the numbers of tests in each series be fixed and the combined number of successes also. It yields the probability that the articles tested in the two series were drawn from the same population--one with a fixed probability of pass. If the total number of successes is not controlled, the results of FET admit of this interpretation only in the limit of large samples. Given that the probability of success could be different in the two populations, it is sometimes claimed that FET can be used to estimate the probability that they differ by prescribed amounts. This claim is unwarranted. The JCS could be faulted for suggesting the test, but they did not underwrite the extended use as in the Army's methodology. (See Kendall and Stuart; also Chapter III).

2) We can know the reliability of an object.

We shall never know the "true" as-manufactured reliability of the components of the Pershing system, and much of such knowledge as we do gain will come at the expense of tactical inventory. It may be that, for the purposes of designing tests of operational reliability, we need not know this a priori probability with any great accuracy; and so methods which treat it as known for this purpose may be satisfactory. This does not justify the assumption when analyzing the results of actual tests.

Section 2. Mathematical Preliminaries

Bayes' Theorem: The Need for a Prior Distribution

Essential to much of what follows is Bayes' Theorem, sketched here as background. The conditional probability of an event B, given that another event A has occurred, is symbolized and defined by

$$P(B|A) = \frac{P(A,B)}{P(A)} \qquad 1.2$$

where P(A) ($\neq$ 0) is the marginal probability of event A, and P(A,B) is the probability of joint occurrence of A and B. One may also speak of P(A/B) = P(A,B)/P(B) with similar meanings and limits, leading to

$$P(B|A)\ P(A) = P(A|B)\ P(B) \qquad 1.3$$

Given that B can occur in n ways Bi (i=1,2,...,n) one of which always occurs with A, we may sum expressions like Eq. 1.3 for the entire set of events Bi

$$P(A)\sum_i P(B_i|A) = \sum_i P(A|B_i)P(B_i) = P(A) \qquad 1.4$$

as the multiplier of P(A) is equal to 1, having encompassed all possible pairings. If P(A) $\neq$ 0, we have Bayes' Theorem:

$$P(B_i|A) = \frac{P(A|B_i)\ P(B_i)}{\sum_i P(A|B_i)P(B_i)} \qquad 1.5$$

Suppose now that events Bi are logically (causally) prior to event A. Then P(Bi) is called the prior distribution of Bi, P(A/Bi) the likelihood of A, given Bi, P(A) the marginal distribution of A, and P(Bi/A) the posterior distribution of Bi. Bayes' Theorem, given in symbols by Eq.1.5, may then be stated in words:

$$\text{Posterior Distribution} = \frac{\text{Prior Distribution X Likelihood (Function)}}{\text{Marginal Distribution}}$$

(This argument holds for both discrete and continuous distributions of probability.)

Likelihood functions are a familiar staple of probability theory, being forecasts of the frequency of chance events A based on presumptions about certain prior events or conditions (a die that is unbiased, the "normal" distribution of errors, half-life of a known radioactive substance). Marginal distributions then are forecasts of

the results of experiments. Bayes' Theorem tells us that inferences about the events Bi which lead to a marginal distribution cannot be derived from the likelihood function alone, but require knowledge of the prior distribution P(Bi) as well. In the context of our task, we need to know more than the results of a set of missile firings to infer the reliability of the missile.

Other requirements of a Bayesian analysis will be discussed as the issues arise.

Section 3. Illustration of an Analysis in Accord with JCS Guidelines

We assume that the missiles and associated ground equipment used in an annual test do come from a homogeneous population, and that the several tests within that year are statistically independent. We assume further that the reliability p is definable, and then may assert that were we to know p, the probability of s_1' successes and f_1' failures in n_1' trials ($n1' = s1' + f1'$) would be by Bernoulli's formula (a likelihood function):

$$\binom{n_1'}{s_1'} p^{s_1'}(1-p)^{f_1'} \quad \text{where} \quad \binom{n}{s} \equiv \frac{n!}{s!\,f!},$$

From component testing, comparison with similar systems, comparison with other products of the same manufacturer, engineering analysis, we should develop an estimate of p and a measure of our confidence in that estimate. Methods exist, e.g. that of Maximum Entropy (Reference 10), for constructing from this information a function with the properties of a probability distribution--a prior distribution. Constraints of reasonableness and mathematical convenience come into the selection process. With limited information at hand, there may be no unique solution. The analyst is free to try several priors and to observe the sensitivity of answers to such variations.

Given a likelihood function, there can generally be found a "conjugate" prior function (so-called because it marries mathematically to the likelihood function); properly a class of such functions, dependent on a limited number of parameters to distinguish members of the class. Conjugate to the Bernoulli's distribution is the Beta distribution, written

$$dP(s_0,f_0) = p^{s_0-1}(1-p)^{f_0-1}\,dp \,/\, B(s_0,f_0) \qquad 1.6$$

$$\text{where} \int_{p=0}^{1} dP(s_0,f_0) = 1, \quad B(s_0,f_0) = \frac{\Gamma(s_0)\Gamma(f_0)}{\Gamma(s_0+f_0)},$$

and $\Gamma(n) = (n-1)!$ for n an integer.

Different sets of the parameters s_o and f_o give rise to functions whose graphs are variously peaked at some locale within the limits of 0 to 1, are relatively flat, are J-shaped and strongly peaked at 0 or 1, or are even U-shaped and strongly peaked at both 0 and 1. It is a rich set of functions.

Taking the product of $dP(s_o,f_o)$ with the Bernoulli function, we get

$$\binom{n_1'}{s_1'} p^{s_1'+s_o-1}(1-p)^{f_1'+f_o-1}dp \,/\, B(s_o,f_o) \qquad 1.7$$

which when integrated over the range of 0 to 1 gives

$$\binom{n_1'}{s_1'} B(s_1,f_1)/B(s_o,f_o) \qquad \text{where} \begin{cases} s_1 = s_o + s_1' \\ f_1 = f_o + f_1' \end{cases}$$

the marginal distribution of s_1' given $B(s_o,f_o)$ as prior. The ratio of Eqs. 1.6 and 1.7 gives the posterior distribution of p for s_1' and f_1' observed:

$$p^{s_1-1}(1-p)^{f_1-1}dp \,/\, B(s_1,f_1) \qquad 1.8$$

explaining my notation and revealing the meaning of conjugation.

From a prior distribution $B(s_o,f_o)$, and a likelihood function for a test of a sample of size n_1', we have created a function which, as a posterior distribution from that experiment, is logically the prior when testing a second sample of size n_2'. This process can be repeated ad libitum, making sample 1 refer to all prior information and sample 2 the latest test.

Now the JCS asks to know the probability that the reliability of sample 2 (and by inference that of the population from which it was drawn) is less than a certain fraction k ($o < k \leq 1$)of the reliability estimate p of sample 1. If the evidentiary basis for this answer lies entirely in the test of n2' items, then we may assume instead a uniform prior distribution, drop the primes on n2', s2', and f2' and represent this probability by

$$P(kp_1) = \int_c^{kp_1} p_2^{s_2-1}(1-p_2)^{f_2-1}\,dp_2 \Big/ B(s_2,f_2)$$

which we then integrate over the distribution of p1 to get the probability that $p2 \leq kp1$:

$$P(p_2 \leq kp_1) = \int_c^{1} p_1^{s_1-1}(1-p_1)^{f_1-1}\,P(kp_1)\,dp_1 \Big/ B(s_1,f_1) \qquad 1.9$$

The probability that p2>kp1 is just 1 minus this result.

As an aid to understanding the generality of this result, consider the case where p1 = r1 x r3 and p2=r2 x r3 where r3 is a reliability factor not subject to degradation but just as much subject to discovery as r1 and r2. Within the framework of Beta-function priors, we might be led to the posterior distribution:

$$dP = K r_1^{s_1-1}(1-r_1)^{f_1-1} r_2^{s_2-1}(1-r_2)^{f_2-1} r_3^{s_3-1}(1-r_3)^{f_3-1} dr_1 dr_2 dr_3 \qquad 1.10$$

where s3(f3) is the total number of observed successes (failures) of the subsystems described by r3. For any values of r3 and k between 0 and 1, P(p2 $\leq$ kp1) = P(r2 $\leq$ kr1). When the latter function is given by integrating Eq.1.10 first over r3 from 0 to 1, it is clear that the result is the same as though r3 = 1 (i.e., it can be ignored). Thus using the criterion p2 $\leq$ kp1 we cam be freed of any concern about reliability factors common to p1 and p2. I would assert that this is a good reason to employ this criterion in preference to the one described next.

The JCS guidance has not always been interpreted as speaking to a proportional reduction in reliability; sometimes it has been interpreted as measuring a reduction of, say, 100d percentage points.*

Instead of Eq. 1.9 we would then use

$$P(p-d) = \int_0^{p-d} p_2^{s_2-1}(1-p_2)^{f_2-1} dp_2 \Big/ B(s_2,f_2)$$

$$\text{and } P(p_2 \leq p-d) = \frac{\int_d^1 p^{s_1-1}(1-p)^{f_1-1}\left\{\int_0^{p-d} p_2^{s_2-1}(1-p_2)^{f_2-1} dp_2\right\} dp}{B(s_2,f_2)\int_d^1 p^{s_1-1}(1-p)^{f_1-1} dp} \qquad 1.11$$

(While we have strayed from the neatness of conjugate functions, by reason of the incomplete integrals, we still have a consistent method. Similar expressions will be found in Reference 8, p. 271.)

* Indeed, the latest revision of the JCS guidance (Reference 5) mandates this form of the criterion.

Eqs. 1.9 and 1.11 give mathematical meaning to the JCS guidance. If at the chosen confidence level it is deemed that there has been no significant change in the reliability between samples 1 and 2, then sample 2 should be merged with sample 1 in preparation for the next year's testing. Other criteria should be examined also (e.g., probability that there has been no significant departure from a nominal value), but that does not refute the translation into mathematics of the JCS guidelines.

At this point I note that much of the historical course of development of mathematics has been devoted to a search for solutions requiring a minimum of actual manipulation of numbers. The approximations used by statisticians are simply good examples of this. The ready availability today of powerful computers reduces the need to employ approximations which may be questionable in particular cases. Most of the calculations to be described here have been carried out on a programmable hand calculator (HP-41) or home computer (Apple, Commodore, etc.). Accordingly, the reader need not be concerned with an apparent intractibility of the formulas. They could be evaluated in the field by the troops of a Pershing fire unit.

There are two matters of concern: the prior distribution and limits to the size of Sample 1. I have already discussed problems with the prior distribution. One assertion made is that with increase in the size of the data base it can become misleadingly narrow, ignoring "unknown-unknowns." A different way of saying this is that tests performed sufficiently long ago may be irrelevant in describing the present state of the missile inventory; the meaning of this argument is that a larger annual test size is needed to compensate for stale data in Sample 1. The question of test size will be the subject of the following chapters. Of course, if there is no evidence of a change in reliability over the years, there is no reason to purge old data.

Section 4. Optimum Test Size

In order to determine the number of missiles which must be procured in the next few years to support a test program through a long period of service life, one must have an estimate of the average annual consumption in testing. To get this estimate, especially if it be glorified by a phase like "optimum test size," one must know what questions the tests are supposed to answer and how frequently. This in turn means "getting into the skull" of the JCS. We must assume that first of all there is sufficient reason to conduct the tests, even at the risk of compromise of properly-classified information. We know that there will be a finite inventory, and that testing reduces that inventory, whether or not it be formally divided into tactical and non-tactical portions. We can then ask the

question: how does the result of an additional test change our perception of the system reliability, and so of the sufficiency of the lesser inventory of missiles to conduct a military mission should it be committed to combat at a future date? Possible answers are discussed in Chapter V. As there are circumstances under which the answer is insensitive to the size of the inventory, we shall spend more time considering the case where inventory for test has no tactical mission.

A long string of heads or tails when flipping pennies is not impossible or even incredible; but after some number, one is entitled to wonder if the coin is biased. Similarly, when testing a missile which is alleged to have high reliability, a string of failures--even a short one--challenges the presumption; contrariwise, a long string of successes tends to be uninformative. In either case there is a practical limit to the value of the additional information in an outcome merely extending such a string.

To address this problem we shall invoke the discipline of Sequential Analysis, to include Sequential Probability Ratio Tests and test series truncation. Much of this is "old hat", having been developed in World War II, most notably by Abraham Wald (Reference 11) working on military problems, and largely standardized by now. It has recently been reported that the methods were independently developed simultaneously by Alan Turing while working at Bletchley Hall to crack the German ENIGMA codes (Reference 12). More importantly there is recent substantive new work not yet "codified" in text books. Two applications of sequential analysis to the Pershing missile test problem will be presented: one by Nozer Singpurwalla and Robert Launer (Chapter III) and one by Michael Woodroofe (Chapter IV). While aspects of the treatment will appear more "frequentist" than Bayesian, both evolve into completely Bayesian solutions. In this paper I shall extract from their work, and comment on it as appropriate. The author of this memorandum is not by profession a statistician, and so requests that the original researchers not be blamed for errors in translating their work into this format.

Chapter III

Launer and Singpurwalla's Proposal

The following submission by Launer and Singpurwalla is the product of over a year of research by the authors, initiated and guided in discussions with the writer of this note. I believe it successfully addresses the problem placed before the authors. Note that all the appendices to this article are to be found at Appendix E.

As the numerical example in the following exposition employs fictitious data and arbitrary values of the parameters α, β, and ∇, the numerical results should not be taken as applicable to the Pershing II problem. The dependencies and the savings from sequential analysis are however clearly indicated, the penalty when tests are batched, and the potential for squeezing information out of small samples. The next chapter reports further steps toward savings through careful test design.

MONITORING THE RELIABILITY OF PERSHING II MISSILES-- A CRITIQUE OF THE CURRENT METHODOLOGY AND A SUGGESTED COMBINED BAYESIAN-SAMPLE THEORETIC APPROACH +

by

Robert Launer*
Nozer D. Singpurwalla**

1. INTRODUCTION, TEST REQUIREMENTS, AND ASSUMPTIONS

The reliability of the Pershing II missile arsenal is an unknown parameter which presumably could change over time. To monitor the reliability, and also to ascertain the amount of change in reliability, if any, a sample of n Pershing II missiles is chosen from the arsenal every year, and each missile fired to observe its success or failure. The testing is destructive, and the arsenal inventory is not replenished. Thus, it is highly desirable to reduce the number of test missiles fired year after year. Also, if possible, it is desirable to have the total number of missiles fired per year be a multiple of three--that is, 3, 6, 9, etc. A stated requirement with respect to the year by year detection of change in reliability is that *a change of* Δ *should be detected with a probability of* π *or more*. Since the test data are

---

+ The authors' appendices are incorporated in this paper as Appendix E. DW

* U.S. Army Research Office, Research Triangle, N.C.
** George Washington University, School of Engineering and Applied Science, Washington, D.C.

of a pass-fail nature, a correct probability model for describing them is the binomial.

Our goal is to determine a sample size and a decision criterion that will satisfy the above requirement, and minimize the total amount of testing. Since each missile is expensive to produce and test, there is a keen desire to incorporate into the analysis all knowledge that is available, both, from the previous tests and engineering experience. Thus a Bayesian point of view is natural here.

2. CRITIQUE OF PRESENT METHODOLOGY

Based on our reading of the pertinent literature that has been made available to us, and our discussions with several people familiar with the test, it is our understanding that the current methodology for analyzing the Pershing II data is based on Fisher's exact test, henceforth FET. We claim that this technique is inappropriate for the situation described above. Furthermore, a modified version of the FET which has been used in similar situations is not appropriate, either. Whereas the FET can be used to detect the equality or otherwise of two binomial populations, it is not designed to detect a specified difference between the two binomial parameters in question. Furthermore, FET does not address the key question of sample size selection, and thus fails to answer the main question posed by our problem. A choice of the sample size should be based on an assumed or target value of the reliability, and this is nowhere apparent in the test.

Given a sample size and the test results from this sample, the FET can give us the "p values" for deciding upon the difference or

otherwise of the two binomial populations in question, and this may be the sole motivation for using this test here.

3. THE COMBINED BAYESIAN-SAMPLE THEORETIC APPROACH PROPOSED HERE

Our proposed approach addresses the issues posed before, and attempts to do this in an economical manner with respect to sample size.

Since reliability changes over time, we introduce an index t, where $t = 1,2,\ldots$; thus $t = 1$ denotes the first year of testing, $t = 2$ denotes the second year of testing, and so on. Let n_t denote the number of missiles to be tested in time period t ; n_t is the (unknown) sample size, one of our decision variables. Let x_t denote the number of missiles that fire successfully in time period t ; note that $0 \leqslant x_t \leqslant n_t$.

Let p_t be the chance that any missile fired at t will fire successfully, or its propensity to do so. Since p_t is unknown to us, we express our uncertainty about it by a probability distribution, say $g(p_t \mid \text{previous failure data, if any, and } H)$. Thus p_t is treated as an unknown parameter, and the vertical line in $g(\cdot)$ denotes conditioned upon or given, and H denotes our background information about p_t . If we have no previous failure data, then $g(p_t \mid H)$ denotes our <u>prior distribution</u> for p_t ; otherwise $g(\cdot \mid \cdot)$ denotes our posterior distribution.

If for each time period t we judge the missiles in the arsenal to be <u>exchangeable</u> (we have here finite exchangeability), then it is appropriate to assume that <u>given</u> p_t , the probability of observing x_t

successful firings in a sample of size n_t is a binomial distribution; that is,

$$P\{x_t \text{ successes in } n_t \text{ firings} \mid p_t\} = \binom{n_t}{x_t} p_t^{x_t} (1 - p_t)^{n_t - x_t} \tag{1}$$

The choice of the sample size n_t is based on the following sample theoretic arguments for testing hypotheses about p_t.

If p_t, the chance that a missile is fired successfully at time t, is large, then the number of failures in a sample of size n_t would tend to be small. Given an n_t and having specified a p_t, let x_t^* be the largest integer for which the chance of observing x_t^* or fewer successes is small, say α; that is,

$$P\{x_t^* \text{ or fewer successes in } n_t \mid p_t\} = \sum_{j=0}^{x_t^*} \binom{n_t}{j} p_t^j (1 - p_t)^{n_t - j} \leqslant \alpha . \tag{2}$$

If p_t were to change to $p_t - \Delta$, with Δ large, then the number of failures in a sample of size n_t would tend to be large; if Δ were small, the number of failures in n_t would tend to be small. Thus, for some small number β,

$$P\{x_t^* \text{ or fewer successes in } n_t \text{ firings} \mid (p_t - \Delta)\}$$

$$= \sum_{j=0}^{x_t^*} \binom{n_t}{j} (p_t - \Delta)^j (1 - p_t + \Delta)^{n_t - j} \geqslant 1 - \beta . \tag{3}$$

If in (2) and (3) we assume that p_t, α, β, and Δ are the only known quantities, then (2) and (3) can be simultaneously solved to obtain an n_t and x_t^*. Once this is done, (2) can be used to test the null hypothesis that the reliability of the missile arsenal at time t

is p_t, with a Type I error α. This is done by accepting (rejecting) the null hypothesis whenever $x_t > (\leq)\ x_t^*$, where x_t is the total number of successfully fired missiles in a sample of size n_t. If $\alpha = .25$ and $\beta = .25$, then (3) assures us that n_t and x_t^* are suitable for detecting the desired changes in reliability. Note that (3) describes the power of the test as specified by (2), for changing values of Δ. If the null hypothesis is accepted, we conclude that the reliability of the missile arsenal at time t is p_t.

In our case p_t is not specified, as it is an unknown parameter which is liable to change over time. What we have instead is

i. a prior distribution for p_t at time $(t-1)$, say
$g(p_t \mid (n_1,x_1), (n_2,x_2), \ldots, (n_{t-1},x_{t-1}), H)$, $t \geq 2$ and
$g(p_1 \mid H)$;

ii. a posterior distribution for p_t at time t, say
$g(p_t \mid (n_1,x_1), \ldots, (n_t,x_t), H)$, for $t \geq 1$.

Thus, if we uncondition on p_t, (2) and (3) would become

$$\int_0^1 \sum_{j=0}^{x_t^*} \binom{n_t}{j} p_t^j (1-p_t)^{n_t-j}\, g(p_t \mid (n_1,x_1), \ldots, (n_{t-1},x_{t-1}), H)\,dp_t \leq \alpha ,$$

for $t \geq 2$, and

$$\int_0^1 \sum_{j=0}^{x_t^*} \binom{n_t}{j} p_t^j (1-p_t)^{n_t-j}\, g(p_1 \mid H)\,dp_1 \leq \alpha , \quad \text{for } t = 1 ; \qquad (4)$$

$$\int_\Delta^1 \sum_{j=0}^{x_t^*} \binom{n_t}{j} (p_t-\Delta)^j (1-p_t+\Delta)^{n_t-j}\, g(p_t \mid (n_1,x_1),\ldots,(n_{t-1},x_{t-1}), H)\,dp_t$$

$$\geq 1 - \beta , \quad \text{for } t \geq 2 ,$$

and

$$\int_{\Delta}^{1} \sum_{j=0}^{x_t^*} \binom{n_t}{j} (p_t-\Delta)^j (1-p_t+\Delta)^{n_t-j} \; g(p_1 \mid H)dp_1 \geq 1 - \beta \;, \quad \text{for} \quad t = 1 \;. \tag{5}$$

In order to obtain the pair (n_t,x_t^*) , for $t \geq 1$, we need to solve (4) and (5) simultaneously. Note that a solution to (4) and (5) would depend on our choice of $g(p_t \mid \cdot)$. If for example, $g(p_t \mid \cdot)$ is a member of the family of beta density functions, then (4) and (5) would involve incomplete beta functions and would call for numerical methods for solving them. A method for undertaking this is described in Appendix A. A computer code for implementing the method of Appendix A is given in Appendix B. An example using the above is in Section 5.

As an alternative to the above, and one which is easy to implement, we replace p_t in (2) and (3) by $\tilde{p}_t$, the modal value of $g(p_t \mid (n_1,x_1), \ldots, (n_{t-1},x_{t-1}), H)$. The modal value is the most likely value of p_t , given all the previous data, and is determined by the prior distribution $g(p_t \mid (n_1,x_1), \ldots, (n_{t-1},x_{t-1}), H)$. The posterior distribution $g(p_t \mid (n_1,x_1), \ldots, (n_t,x_t), H)$ represents our best assessment of the arsenal reliability at time t , given all the data up to and including that obtained at t . Its model value $\hat{p}_t$ could be used as a single number which describes p_t . In the next section, we discuss an implementation of the above alternative procedure. An implementation of the main procedure follows along similar lines, with the exception that in computing the pair (n_t,x_t^*) p_t is not replaced by the modal value of its prior distribution.

3.1 Assessing Our Uncertainty about p_t and Procedure Implementation

Since p_t can take values between 0 and 1, a convenient but flexible way for us to express our uncertainty about p_t is via the family of beta density functions on (0,1). Thus,

1. We start off our assessment and monitoring procedure by assigning a prior distribution for p_1, say $g(p_1 \mid \gamma,\delta,H)$, which for the two unknown parameters $\gamma > 0$ and $\delta > 0$ is a beta density function

$$g(p_1 \mid \gamma,\delta,H) = \frac{\Gamma(\gamma+\delta)}{\Gamma(\gamma)\Gamma(\delta)} p_1^{\gamma-1} (1-p_1)^{\delta-1} , \quad 0 < p_1 < 1 . \tag{6}$$

   The modal value of the above density is

$$\tilde{p}_1 = \frac{\gamma-1}{\gamma+\delta-2} .$$

   Clearly, p_1 best describes in the form of a single number our assessment of $\tilde{p}_1$, prior to testing at time $t = 1$. Furthermore, $\tilde{p}_1$ is also to be used for determining the pair n_1 and x_1^*, for testing at time $t = 1$.

2. We thus replace p_t by $\tilde{p}_1$ in (2) and (3), and simultaneously solve these to obtain n_1 and x_1^*. [In Appendix A we discuss how to obtain n_1 and x_1^* without using $\tilde{p}_1$, and by directly solving (4) and (5).]

3. We take a sample of size n_1 and test these to determine x_1, the number of missiles that fire successfully. If $x_1 >(<)\ x_1^*$, we accept (reject) the hypothesis that the reliability of the missile arsenal at time 1 is $\tilde{p}_1$.

4. If we accept the above hypothesis, then we update our opinions

about p_1 in light of n_1 and x_1 via the posterior distribution $g(p_1 \mid (n_1, x_1), H)$. The modal value of this posterior distribution is

$$\hat{p}_1 = \frac{\gamma + x_1 - 1}{\gamma + \delta + n_1 - 2} ,$$

and this number best summarizes our assessment of p_1 after testing at time 1. We now go to step 5.

5. If the aforementioned hypothesis is rejected, our choice of γ and δ needs to be revised. This should be done following a more detailed analysis about p_1 . We then go back to stage 1.

6. The posterior distribution $g(p_1 \mid (n_1, x_1), H)$ now serves as the prior distribution for p_2 , and its modal value $\hat{p}_1$ is set equal to $\tilde{p}_2$. Thus

$$\tilde{p}_2 = \frac{\gamma + x_1 - 1}{\gamma + \delta + n_1 - 2} ,$$

and p_t is now replaced by $\tilde{p}_2$ in (2) and (3), which are solved for n_2 and x_2^* . [In Appendix A we discuss how to obtain n_2 and x_2^* by directly solving (4) and (5).]

7. We now repeat the steps 3 through 6, and continue the above procedure. Thus, at time (t-1) we have

$$\hat{p}_{t-1} = \tilde{p}_t = \frac{\gamma + x_1 + x_2 + \ldots + x_{t-1}}{\gamma + \delta + n_1 + n_2 + \ldots + n_{t-1} 2} \qquad (7)$$

as our single best assessment of the reliability of the arsenal at time (t-1), after observing the results of the test at

time $(t-1)$. It also represents our choice for p_t in equations (2) and (3), for determining the sample size n_t and the decision variable x_t^*.

8. Suppose that at time t, we test n_t items, observe x_t successes, and based on this result, reject the null hypothesis that $p_t = \tilde{p}_t = \hat{p}_{t-1}$. Then we conclude that the reliability of the arsenal has changed from its previous value $\hat{p}_{t-1}$. When this happens, we investigate the cause for this change, choose some new values, say γ' and δ', and estimate p_t by

$$\hat{p}_t = \frac{\gamma'+x_t-1}{\gamma'+\delta'+n_t-2} .$$

We now continue as before, bearing in mind that the previous date $(n_1,x_1), \ldots, (n_{t-1},x_{t-1})$ are no more appropriate for inclusion in our assessment process.

An alternative to the beta prior which has properties of robustness is currently under investigation. However, there is no assurance that the alternative prior will be void of computational difficulties.

3.2 Sequential Sampling to Reduce the Amount of Testing

At any stage t, given an n_t and x_t^*, a further reduction in the amount of missiles tested can be achieved if the testing is done sequentially, one item at a time. Specifically, we would test one item at a time; and stop the test as soon as x_t the number of successes is larger than x_t^*. Thus, ideally, the number of missiles tested could be

as few as $x_t^* + 1$; this implies a saving of $n_t - x_t^* - 1$. The maximum of missiles tested would of course be no greater than n_t . The resulting sample size, that is the number of missiles actually tested at each stage is known as a <u>curtailed sample</u>.

For the above scheme, given p_t we can compute $E(n_t|p_t)$ the expected number of missiles tested using standard arguments--these are shown later. However, since p_t is not known, we average out p_t with respect to its prior distribution to obtain $E(n_t)$, the unconditional expectation of the number of missiles tested at each stage under the sequentially taken curtailed sample. This is shown below.

Given n_t and x_t^* , the probability that $n_t = x$, when a sequential sampling scheme is used is

$$p\{n_t=x|p_t\} = \begin{cases} \binom{x-1}{n_t-x_t^*-1} (1-p_t)^{n_t-x_t^*} \, p_t^{x-(n_t-x_t^*)} , & n_t - x_t^* \leqq x \leqq x_t^* \\ \binom{x-1}{n_t-x_t^*-1} (1-p_t)^{n_t-x_t^*} \, p_t^{x-(n_t-x_t^*)} & \\ \quad + \binom{x-1}{x-x_t^*-1} (1-p_t)^{x-x_t^*-1} \, p_t^{x_t^*+1} , & x_t^* < x \leqq n_t . \end{cases}$$

In order to obtain $P\{n_t=x\}$, we average out the above by $g(p_t|\cdot)$, where

$$g(p_t|\cdot) = \frac{\Gamma(\gamma+\delta)}{\Gamma(\gamma)\Gamma(\delta)} \, p_t^{\gamma-1}(1-p_t)^{\delta-1} .$$

When the above is done, we have

$$
p[n_t=x] = \begin{cases} \dbinom{x-1}{n_t-x_t^*-1} \dfrac{\Gamma(\gamma+\delta)}{\Gamma(\gamma)\Gamma(\delta)} \dfrac{\Gamma(x-n_t+x_t^*+\gamma)\Gamma(n_t-x_t^*+\delta)}{\Gamma(\gamma+\delta+x)} & \text{for } n_t-x_t^* \leq x \leq x_t^* \\[2ex] \dbinom{x-1}{n_t-x_t^*-1} \dfrac{\Gamma(\gamma+\delta)}{\Gamma(\gamma)\Gamma(\delta)} \dfrac{\Gamma(x-n_t+x_t^*+\gamma)\ (n_t-x_t^*+\delta)}{\Gamma(\gamma+\delta+x)} & \\ \quad + \dbinom{x-1}{x-x_t^*-1} \dfrac{\Gamma(\gamma+\delta)}{\Gamma(\gamma)\Gamma(\delta)} \dfrac{(x_t^*+1+\gamma)\Gamma(x-x_t^*-1+\delta)}{\Gamma(\gamma+\delta+x)} & \text{for } x_t^* < x \leq n_t , \end{cases}
$$

from which $E(n_t)$ can be computed. The above formula can also be used to plot a histogram of the various values of n_t, for each stage t.

If the sequential tests are to be done in batches of 3 rather than testing a single item at a time, the savings in the number of items tested will be less. However, this is still better than compulsarily testing all the n_t items. We do not have a general formula like (9) above to compute the expected sample size. The calculations will have to be done on an enumerative basis. These are shown in Appendix C.

4. COMMENTS ON THE PROPOSED APPROACH

The proposed approach is a combination of sample theory and Bayesian statistics. The former is used to determine the sample size, and the latter is used for inference about p_t. One may express reservations about a procedure in which two philosophical viewpoints are used simultaneously. However, upon closer examination of the approach, such a concern should be dispelled, since the sample theory approach is not used for making inferences about p_t ; it is used for choosing a sample size. The selection of the sample size after averaging out p_t with respect to its distribution $g(p_t \mid \cdot)$, see equations (4) and (5), makes our analysis fall under the category of what is known as pre-posterior analysis, a perfectly legitimate device within the Bayesian paradigm [cf. Box (1982)].

The monitoring of p_t is done within the Bayesian framework, and besides "coherence" it has the advantage of inducing economy by virtue of the fact that all our relevant previous data are incorporated into the analysis. Furthermore, it allows the incorporation of any engineering or judgmental knowledge that we may have about the missiles into our analysis -- this is done via the parameters γ and δ or γ' and δ' , etc.

5. APPLICATIONS TO DATA

Our proposed approach is designed to specify a sample size for testing at each stage, and thus its effectiveness cannot be fully appreciated if we apply it to existing data. However, we shall apply it

to some given (sanitized) success failure data to demonstrate the fact that the computations of Appendix A can be undertaken, and to compare the results of our main procedure and the simplified alternative, described in Section 3.1. In Table 1, we present the given success failure data, our Bayesian estimate of the mode of p_t at each stage using a uniform prior distribution at stage 0 updated at successive stages using failure data, and the values of x_t^* and N_t using the main procedure and the alternative.

A few facts emerge from an examination of Table 1.

1. A large number of items to be tested is called for, when the prior is uniform, with mode .5 .
2. The number of items to be tested is the smallest when the mode of p_t is closest to 1, namely, at .9 .
3. The number of items to be tested under the main procedure is always equal to or larger than that under the alternate procedure. This is because the alternate procedure puts all the probability mass at the mode, whereas the main procedure disperses the probability mass over [0,1] , with a concentration at the mode.

5.1 Results of Curtailed Sequential Sampling

The sequential sampling approach discussed in Section 3.2 was applied to the data and the results of Table 1. The n_t and the x_t^* values considered were those given by the "alternative procedure"; this procedure gave us smaller values of the n_t's than the main procedure.

TABLE 1

Results for Main Procedure and Alternative, Using Sanitized Data, and Assuming a Uniform Prior at Stage 0

Stage t	Data		Mode of p_t	Computed Values of x_t^* and n_t			
				Main Procedure		Alt. Procedure	
	Success	Failure		x_t^*	n_t	x_t^*	n_t
0			.500	2	29	5	17
1	6	0	.875	8	13	9	13
2	11	1	.900	10	14	8	11
3	11	1	.906	11	15	8	11
4	11	1	.909	8	11	8	11
5	9	3	.875	9	13	9	13
6	9	3	.853	10	15	8	12
7	8	4	.825	9	14	9	14
8	4	0	.833	11	17	9	14
9	3	2	.820	10	16	9	14
10	9	0	.837	10	15	9	14
11	8	1	.841	10	15	10	15
12	7	2	.836	10	15	9	14
13	9	0	.848	10	15	8	12
14	7	1	.850	10	15	8	12

The expected sample sizes when testing is sequential, in batches of 3 as well as one item at a time, were computed. These are shown in Table 2. The advantage of testing one item at a time is clear from an inspection of columns 2 and 3 of Table 2.

We also note the overall reduction in sample size using the approach of this paper. The expected sample size can be as small as 9.

The detailed calculations leading us to Columns 2 and 3 of Table 2 are given in Appendix C.

6. PROPOSED FUTURE WORK

An objectionable feature of the proposed procedure, from a Bayesian point of view, is the testing of hypotheses about $\tilde{p}_t$ using the decision variables x_t^*, $t = 1,2,\ldots$. The proper Bayesian way to study this problem would be via a Kalman filter model which contains two unknown states of nature, p_t and m_t, where m_t denotes the drift in p_t. The Kalman filter would not only have the ability to monitor the reliability of the arsenal, but would also provide us with a vehicle for predicting the future arsenal reliability. The following are our ideas on how a Kalman filter model for this problem can be developed.

Let Y_t denote some transform of x_t/n_t, and one which makes Y_t approximately normal. The observation equation for the Kalman filter model would be

$$Y_t = p_t + \gamma_{1t}$$

where γ_{1t} is a disturbance term with mean 0 and variance σ_{1t}^2.

We can postulate the following as system equations:

$$p_t = m_t + \gamma_{2t}, \quad \text{and}$$

$$m_t = m_{t-1} + \gamma_{3t} .$$

TABLE 2

Expected Sample Size for Curtailed Sequential Sampling in Batches of Size 3 and Size 1.

Stage t	Expected Sample Size for Batch Size 3	Expected Sample Size for Batch Size 1	x_t^*	n_t
0	11.84	10.91	5	17
1	12.03	10.66	9	13
2	10.29	9.45	8	11
3	10.37	9.51	8	11
4	10.40	9.54	8	11
5	12.28	11.08	9	13
6	11.07	10.16	8	12
7	12.84	11.74	9	14
8	12.79	11.69	9	14
9	12.87	11.78	9	14
10	12.78	11.67	9	14
11	13.59	12.72	10	15
12	12.78	11.68	9	14
13	11.14	10.22	8	12
14	11.14	10.21	8	12

In the above equations, we are saying that p_t , the unknown state of nature, consists of a low frequency drift term m_t , which represents a smooth variation in p_t , and γ_{2t} , which is a high frequency component that represents drastic changes in p_t . We assume that γ_{2t} is a normal variate with mean 0 and variance σ^2_{2t} . The drift term is assumed constant, except for slight disturbances in it; these are described by γ_{3t} , which is also assumed normal with mean 0 and variance σ^2_{3t} .

The Kalman filter solution would result in uncertainty statements about p_t and m_t , via their distribution functions. These, of course, would be conditioned on $(n_1,x_1), \ldots, (n_t,x_t)$. Large values of m_t would indicate a drift in the arsenal reliability, and so m_t could be used to monitor the change in the arsenal reliability.

It appears that the Kalman filter solution would have several advantages over the proposed approach. The problem of choosing n_t in the context of a Kalman filter is an open question, and this calls for some basic research, assuming that this has not been done before.

A third possible direction for future research is the development of a sequential procedure for testing the missiles. A sequential procedure employing Bayesian considerations may add a further dimension to this problem.

Chapter IV
Woodroofe's Proposal

The proposals of Michael Woodroofe are not yet formally documented, but are contained in a series of letters and lecture notes (References 13-17). In this chapter I shall mostly quote from this material with the author's permission, noting that any published versions may differ markedly from those given here. I accept responsibility, however, for the accuracy of the material quoted and the interpretations and extensions of it.

All of the calculations described in this chapter were carried out by Dr. Woodroofe and/or myself. I have programmed most of them for an HP-41C, and listings are given in Appendix D. Instructions and copies on magnetic cards are available. Dr. Woodroofe has used an Apple computer.

<u>Section 1</u>. (Extract from Reference 14).

The Truncated Sequential Probability Ratio Test.

Illustration with a sequential test of the type of savings which are possible and the loss of information which results from the savings. Note that the process starts with the conventional Uniformly Most Powerful test, to be terminated when a specific number Sn of failures has been observed; or when, out of a planned test of size n, the number of observed successes assures that the number of failures cannot reach Sn; or after n tests if not terminated earlier. The choice of n is at this time arbitrary; the value 12 was used in the example to permit comparison to the Pershing test program, past and planned.

/We start with a discussion of/ the problem of sequentially testing /such that/ that a failure probability does not exceed a given level. I will illustrate the type of savings which are possible and the loss of information which result from the savings with a specific example.

Let $X_1, \ldots X_{12}$ be i.i.d.* random variables which take the values 1 and 0 with probabilities p and $q = 1-p$, where $0 < p < 1$, is unknown; and consider the problem of testing

$$H_o: \ p \leq .15 .$$

Let $S_k = X_1 + \ldots + X_k$, $\qquad 1 \leq k \leq 12$.

Then the (UMP)** test which rejects H_o if and only if $S_{12} \geq 4$ has power function

$$(1) \quad \beta_o(p) = 1 - \sum_{k=0}^{3} \binom{12}{k} p^k q^{n-k}, \qquad 0 < p < 1 .$$

Of course, it may not be necessary to take all 12 observations to determine whether $S_{12} \geq 4$. The test may be curtailed at time

$$t_o = \min\{k \geq 1: \ S_k \geq 4 \ \text{ or } \ S_k \leq k-9\} .$$

Then

$$(2) \quad E_p(t_o) = \sum_{k=4}^{12} k\binom{k-1}{3} p^4 q^{k-4}$$

$$+ \sum_{k=9}^{12} k\binom{k-1}{8} q^9 p^{k-9}, \quad 0 < p < 1$$

---

* Identically and Independently Distributed.
** Uniformly Most Powerful.

is the expected sample size of the curtailed test.

Selected values of $\beta_o(p)$ and $E_p(t_o)$ are listed in columns 2 and 4 of Table 1 below.

Observe that the type I error probability is .0922 when $p = .15$ and the type II error probability is .2253 when $p = .4$.

I tried to construct a truncated version of the SPRT whose power function matched β_o as closely as possible. Wald's approximations allow one to match the power function at two points. I picked .15 and .40. Wald's approximations then give formulas for the upper and lower stopping boundaries in the (k, S_k) plane. These are listed in columns 2 and 3 of Table 2. There are two problems with these boundaries: Wald's approximations tend to overestimate the error probabilities; and I wanted the test to take at most 12 observations. After some experimentation with formulas (3) and (4) below, I was led to the upper and lower boundaries listed in columns 4 and 5 of Table 2.

Thus, I considered the sequential test which takes

$$t = \min\{k \geq 1: \ S_k \leq a_k \ \text{ or } \ S_k \geq b_k\}$$

observations and rejects H_o if and only if $S_t \geq b_t$, where a_k and b_k are as in Table 2.

The power function and expected sample size may be easily computed. Let

$$f_k(j,p) = P_p\{S_k = j \ , \ t > k\}$$

for $k = 0, \ldots, 11$, $j = 0, 1, 2, \ldots,$ and $0 < p < 1$. Then the power function and expected sample size are

$$\beta(p) = \sum_{k=1}^{11} f_{k-1}(b_k - 1, p) \cdot p \tag{3}$$

and

$$E_p(t) = \sum_{k=1}^{12} k\{f_{k-1}(b_k - 1, p)\ p + f_{k-1}(a_k, p) q\} \tag{4}$$

for $0 < p < 1$. Thus, one need only compute the values of f_k; and this is easy in view of the initial conditions, $f_0(0,p) = 1$ and $f_0(j,p) = 0$ for $j \neq 0$, and the recursion

$$f_k(j,p) = [p\ f_{k-1}(j-1,p) + q\ f_{k-1}(j,p)]\ I\{a_k < j < b_k\} \tag{5}$$

for $k = 1, \ldots, 12$, $j = 0, 1, 2, \ldots$, and $0 < p < 1$. Here I_A denotes the indicator of A.

The power function and expected sample size may be computed from (3), (4), and (5). Selected values are listed in columns 3 and 5 of Table 1.

Observe that the power functions β_o and β differ by at most .0103 for the values computed. This is much better than I had expected when I began the exercise. Observe also that the expected sample size of the modified SPRT is substantially smaller than that of the curtailed test when p is small.

After the test has been performed, one may set confidence limits for p by using the relationship between tests and confidence intervals. Order the possible outcomes in a clockwise manner, as in column 1 of Table 3. For each r, $0 < r < 1$, one may test the hypothesis

$$K_r: \quad p \geq r$$

as follows: the acceptance region $A(r)$ of the test consists of an initial segment of outcomes, in the order of Table 3; one includes precisely enough outcomes to make

$$P_r(A(r)) \geq .90 .$$

Then, after the test has been performed, an upper confidence bound p^* for p may be obtained from the relation

$$p \leq p^* \quad \text{iff} \quad (t, S_t) \in A(p) .$$

This is essentially the approach of Siegmund (1978, Biometrika), but substitutes exact calculations for his approximations.

I list some approximate 75% upper confidence bounds for p in Table 3. These were obtained by linear interpolation with formulas like (3).

To the extent that the modified sequential test takes fewer observations than the curtailed test, one may expect less accurate estimation of p.

Table 1: Power Functions and Expected Sample Sizes

p	$\beta_o(p)$	$\beta(p)$	$E_p(t_o)$	$E_p(t)$
.05	.0022	.0022	9.47	6.93
.10	.0256	.0251	9.92	7.85
.15	.0922	.0899	10.23	8.62
.20	.2054	.2004	10.40	9.13
.25	.3512	.3434	10.31	9.35
.30	.5075	.4975	10.02	9.30
.40	.7747	.7644	9.00	8.57
.50	.9270	.9204	7.77	7.42

Here: Column 1 is computed from (1), column 2 from (3), column 3 from (2), and column 4 from 4.

Table 2: Upper and Lower Stopping Boundaries in the (k, S_k) Plane

	The SPRT		Modified	
k	a_k^*	b_k^*	a_k	b_k
1	-1	2	-1	3
2	-1	3	-1	3
3	-1	3	-1	3
4	-1	3	-1	4
5	0	3	-1	4
6	0	4	0	4
7	0	4	0	4
8	1	4	0	4
9	1	4	1	4
10	1	5	1	4
11	1	5	2	4
12	2	5	3	4

Here columns 2 and 3 are from Wald's approximations; columns 4 and 5 are ad hoc approximations.

Table 3: Approximate 75% Upper Confidence Bounds

Outcome t	S_t	Confidence Bound
3	3	
5	4	.91
6	4	.70
7	4	.61
8	4	.55
9	4	. 5
10	4	.45
11	4	.42
12	4	.39
12	3	.34
11	2	.29
9	1	.21
6	0	

Comment by DW:

As indicated in Chapter III, expectations of ρ and E can be computed based on a prior probability distribution. Closed-form solutions exist for ρo and Ep(to) for a Beta prior, among others. For $\rho(p)$, and Ep(t), numerical integration is necessary. Other indices derived from the fk(j,p) in manners like that for ρ or E(t) can also be meaningfully be averaged over a prior distribution. As Ep(t) has here a narrow range of variation, its expectation value will not be very sensitive to the choice of the prior distribution.

Section 2. (Extract from Reference 15).

To clarify some of the points raised in Section I, Woodroofe provided a more extensive treatment of the development of the limits on observed successes and failures at which the test is terminated. It begins with the method described by Wald (op. cit.) and then continues with a procedure, somewhat judgmental, for modifying those boundaries to reduce the expected size of the test while retaining its power.

1) Testing H_O: $\theta > .15$ is the same as testing $\theta' = 1 - \theta < .85$. If you want to have

$$P_\theta\{\text{decide } \theta' > .85\} < \alpha_0 \qquad \text{for } \theta' < .85$$

and

$$P_\theta\{\text{decide } \theta' < .85\} < \alpha_1 \qquad \text{for all } \theta' > \theta'_1 > .85,$$

where α_0 and α_1 are small and $.85 < \theta'_1 < 1$, then you cannot simply reverse the roles of zero and 1 in the test described in my earlier letter. A new test must be constructed. See (2) below.

In /Section I/ θ was the probability of a system failure.

2) For testing H_O: $\theta < \theta_0$ at level α_0 with type II error at most α_1 when $\theta > \theta_1$, where $0 < \theta_0 < \theta_1 < 1$ are specified, the SPRT continues sampling as long as

$$1/A < L_n < B \qquad (*)$$

where $B = (1-\alpha_1)/\alpha_0$, $A = (1-\alpha_0)/\alpha_1$, and L_n is the likelihood ratio. One finds

$$L_n = \exp\{\Delta_1 S_n - n\,\Delta_0\}$$

where

$$\Delta_1 = \log \theta_1(1-\theta_0) - \log \theta_0(1-\theta_1)$$

$$\Delta_0 = \log(1-\theta_0) - \log(1-\theta_1)$$

and $$S_n = X_1 + \ldots + X_n, \qquad n > 1.$$

Since S_n are integer valued, equation (*) may be rewritten

$$a_n' < S_n < b_n'$$

$$a_n' = \left[\frac{1}{\Delta_1}(n\Delta_0 - \log A)\right]$$

$$b_n' = \left[\frac{1}{\Delta_1}(n\Delta_0 + \log B)\right] + 1$$

where [x] is the greatest integer which is less than or equal to x.

Suppose now that one wants the test to be truncated at M say. Then one wants boundaries a_n and b_n, $1 < n < M$. What I did in the example was the following. Let a_M and b_M be such that

$$a_M' < a_M = b_M - 1 \text{ and } b_M < b_M',$$

say two integers near the middle of the interval from a_M' to b_M'. Then let

$$a_n = \max\{a_n', a_M - (M-n)\}$$

and $$b_n = \min\{b_n', b_M\}$$

for $n < M$. This gives a first approximation to the boundary. In the example, I then computed the power function of the sequential test with boundaries a_n and b_n and compared it with the power function of the fixed sample size test. I then changed a few of the boundary points to get better agreement between the two power functions. The adjustments were minor and tended to make the continuation region fatter.

The reason that you can't pin me down on the adjustments is that it is a trial and error operation.

(3) In the example,

$$P_\theta\{t=k, S_k = b_k\} = f_{k-1}(b_k - 1;\theta)\cdot \theta$$

and

$$P_\theta\{t=k, S_k = a_k\} = f_{k-1}(a_k;\theta)\cdot (1-\theta)$$

Then $P_\theta\{\bar{X}_t > x\}$ is the sum of these probabilities over all pairs (k, a_k) and (k, b_k) for which $a_k/k > x$ or $b_k/k > x$.

4) For inverse sampling there is just one boundary. For curtailed sampling, there are two. Let

$$t^+ = \min\{k > 1: \quad S_k > 4\}$$

and

$$t^- = \min\{k > 1: \quad k - S_k > 9\}$$

Then $\qquad E_\theta(t^+) = 4/\theta$

and $\qquad E\theta(t^-) = 9/(1-\theta)$

The stopping time for the curtailed fixed sample size test is

$$t_0 = \min(t^+, t^-)$$

So $\qquad E_\theta(t_0) < \min\{E_\theta(t^+), E_\theta(t^-)\}$

When $\theta = .15$, $E_\theta(t^-) = 10.6$.

The formulas for $E_\theta(t^+)$ and $E_\theta(t^-)$ hold for all θ, $0 < \theta < 1$.

5) I think of the boundaries as a modified S.P.R.T. In the example, they were similar to the curtailed fixed sample size test, but sufficiently different to reduce the expected sample size by about 1 over the range of interest.

6) The calculations in my letter to Launer are for fixed θ. To do a Bayesian calculation, one would average them over θ values

The formulas which I gave for computing the power and expected sample implicitly assume that that the boundaries a_n and b_n are non-decreasing in n.

Section 3 (Extract from Reference 16).

The Truncated SPRT, Aggregated over Several Tests.

Derivation of a conservative estimate of the probability that in 10 years of testing, at 12 missiles planned for expenditure each year, no more than, say 100, will be needed using the proposed stopping rules.

This is to explain how savings in expected sample size may be translated into savings of units which must be purchased prior to the experimentation. For definiteness, I illustrate the method with the truncated SPRT, which is described in / Section I/

In particular, recall the computation of

$$f(k,j;p) = PR(T>k, S_k=j),$$

where p denotes the true failure probability, S_k denotes the number of failures after k units have been tested, and t denotes the stopping time. From this, one gets

$$G(k;p) = Pr(T<k) = 1 - \Sigma_{j=0}^{k} f(k,j;p)$$

and $\quad g(k;p) = Pr(T=k) = G(k;p) - G(k-1;p)$

for $k = 1,\ldots,12$ and $0 < p < 1$.

Suppose that the truncated SPRT is run n times, say once each year for n years, where n is a positive integer. Then there will be a sequence $p_1,\ldots,p_n$ of unobservable true failure probabilities and a sequence $t_1,\ldots,t_n$ of random sample sizes. Here I regard $p_1,\ldots,p_n$ as unknown parameters, and suppose that $t_1,\ldots,t_n$ are independent random variables for which

$$Pr(t_i=k) = g(k;p_i)$$

for $k = 1,\ldots,12$ and $i = 1,\ldots,n$. If $p_1,\ldots,p_n$ are really random variables, then the calculations described below are valid, if the conditional distribution of $t_1,\ldots,t_n$ given $p_1,\ldots,p_n$ is as just described.

Let T denote the total number of units used during the tests,

$$T = t_1 + \ldots + t_n.$$

Then the distribution of T is required. The distribution of T is the convolution of the individual distributions of $t_1, \ldots, t_n$. This depends on $p_1, \ldots, p_n$ in a complicated manner, but it is possible to find the sharp bound which is valid for all possible choices of $p_1, \ldots, p_n$. That is, it is possible to find a function H for which

$$\Pr(T < k) > H(k)$$

for all $k = 1, \ldots, 12n$ and all possible choices of $p_1, \ldots, p_n$.

I describe the derivation below.
The values of H are included in Table 2 in the special case that $n = 10$. Observe that then

$$\Pr(T > 105) < .054$$

for all $p_1, \ldots, p_n$. The bound is reasonably sharp, since $\Pr(T > 105) = .050$ when all of $p_1, \ldots, p_n$ are equal to .27.

While the bound is sharp, the approach is conservative, since it ignores data from previous years and assumes the worst possible values for $p_1, \ldots, p_n$. If an independent verification is required for each year, then some of this conservatism may be unavoidable.

The derivation of the bound uses the notion of stochastic dominance. If X and Y are random variables with distribution functions F and G, then Y is said to be stochastically larger than X if and only if $G(z) < F(z)$ for all z. If X and X' are independent random variables and Y and Y' are independent random variables and if Y and Y' are individually stochastically larger ~~and~~ X and X', then $Y+Y'$ is stochastically larger than $X+X'$ (as is easily verified); and this result extends from two summands to several. To apply this result, let

$$G(k) = \min G(k;p),$$

where the minimum extends over $0 < p < 1$. Then, for any choice of $p_1, \ldots, p_n$, the distribution of T is stochastically dominated by the sum of n independent random variables having common distribution function G. Computing G is straightforward. For $k < 6$, the minimum is attained when $p = 0$ and $G(k) = 0$. For $k > 6$, I computed $G(k;p)$ for a grid of p values and took the minimum over this grid. The values are listed in Table 1. I used a grid width of .01.

TABLE 1. Values of G(k;p)

p	k =	6	7	8	9	10	11
.24		.2313	.2590	.2966	.5032	.5556	.7723
.25		.2222	.2535	.2955	.4967	.5537	.7685
.26		.2144	.2496	.2962	.4923	.5538	.7661
.27		.2081	.2474	.2987	.4900	.5559	.7651
.28		.2032	.2468	.3030	.4897	.5599	.7654
.29		.1996	.2477	.3088	.4914	.5657	.7669
.30		.1974	.2501	.3163	.4949	.5731	.7696
.31		.1964	.2540	.3252	.5002	.5819	.7735
.32		.1967	.2593	.3355	.5072	.5922	.7783
.33		.1981	.2659	.3472	.5156	.6036	.7840
.34							
.35							

	a	b	c	d	e	f	g
Minimum	.1964	.2468	.2955	.4897	.5537	.7651	1.0

Mean and St dev $\mu = 9.4528$ $\sigma = 2.1992$

$= 12\cdot 1 - (a+b+c+d+e+f)$

$\sigma^2 = 12^2 - (12+11)f - (11+10)e - (10+9)d - (9+8)c - (8+7)b - (7+6)a$

Notes: $G(12;p) = 1$ for all $0 < p < 1$; the minimum is zero for $k < 5$; μ and σ are the standard deviation of the minimizing distribution.

TABLE 2. Values of H

k	1 - H(k)	H(k) - H(k-1)
100	.2026	.0460
101	.1622	.0404
102	.1273	.0349
103	.0978	.0295
104	.0734	.0244
105	.0537	.0197
106	.0382	.0155
107	.0263	.0118
108	.0175	.0088
109	.0112	.0063
110	.0069	.0043
111	.0040	.0029
112	.0022	.0018
113	.0012	.0011
114	.0006	.0006
115	.0002	.0003

Comments by DW:

Let $g(k) = G(k)-G(k-1)$. 4.1

Then $$d(n,z) \equiv \sum_{k=0}^{n} z^{k} g(n-k)$$ 4.2

is a generating function of the distribution g(k). The generating function for the dominant of m years' test results is then

$$D(n,m) \equiv \left[d(n,z)\right]^{m} = \sum_{J=0}^{nm} z^{J} d_{J} \text{, say,}$$ 4.3

and the dominant of the probability that a specific number J of tests can be forgone is given by the coefficient dJ of zJ in the expansion of D(n,m).

In our example n=12, and the g(k) for k < 6 are all zeros. Sample data are given in Table 3. So, for m=10,

$$D(12,10) =$$ 4.4

$$\left[g(12) + z g(11) + z^{2} g(10) + z^{3} g(9) + z^{4} g(8) + z^{5} g(7) + z^{6} g(6)\right]^{10}$$

$$= \left[g(12)\right]^{10} + 10 z\, g(11)\left[g(12)\right]^{9} + \cdots .$$

TABLE 3

g(k)

	P = .85		P = .75	
	Batch Size		Batch Size	
k	1	3	1	3
12	.2349	.5103	.0940	.4433
11	.2114	0	.1258	0
10	.0640	0	.2235	0
9	.1942	.2933	.1694	.3604
8	.0487	0	.0361	0
7	.0504	0	.1549	0
6	.1964	.1964	.1963	.1963

In Woodroofe's notation

$$d_J = H(nm-J) - H(nm-J-1).$$

In particular, in our case,

$$d_0 = H(120) - H(119) = 1 - H(119) = [g(12)]^{10}$$

is the dominant of the probability that all 120 are required (none can be foregone). It follows that

$$e_J \equiv \sum_0^{J} d_i = 1 - H(nm-1-J)$$

is the dominant of the probability that at most J can be forgone; the generating function for eJ is

$$E(n,m) = \sum_0^{nm} z^J e_J = D(n,m)/(1-z).$$

The calculation of the dJ or eJ presents no difficulty except possibly in the control of round-off errors for J large. Sample results are given in Tables 4 and 5 partly repeating material in Table 2, with differences presumably due to differences in accuracy between our computers.

In actual conduct of Follow-on Tests, three failures in a row, or two with an identifiable cause, would be sufficient justification for halting the test until the problem were (identified and) fixed. There would then remain some number of missiles from that year's allocation available for intensive investigation of the fault and for demonstration of remediation. It is not clear that any additional missiles would need to be allocated to those missions, as they could serve the FOT mission at the same time.

It is a trivial matter to revise the expression for D(n,m) to treat the case of batched tests: for example, in groups of 3. Tables 3-5 compare the results for single and triple tests. For the data in the example, whatever the number of missiles considered an adequate inventory for 10 years' testing without batching, about 6-10 more would be required when fired in batches of 3. The analysis in Chapter III gave a similar result.

Up to this point the development has assumed that up to 12 would, in fact, be expended if necessary to provide the foundation for an annual confidence estimate. The question now is: why

TABLE 4

P = .85

Singles			Batches of 3		
k	dJ= H(k)-H(k-1)	eJ= 1-H(k)	dj	eJ	J
120	5.1E-7	5.1E-7	.0012	.0012	0
119	4.5E-6	5.1E-6			1
118	2.0E-5	2.5E-5			2
117	.0001	.0001	.0069	.0081	3
116	.0001	.0002			4
115	.0003	.0006			5
114	.0006	.0012	.0224	.0305	6
113	.0011	.0022			7
112	.0018	.0040			8
111	.0029	.0069	.0511	.0816	9
110	.0043	.0112			10
109	.0063	.0175			11
108	.0088	.0263	.0902	.1718	12
107	.0118	.0382			13
106	.0155	.0536			14
105	.0196	.0733	.1291	.3010	15
104	.0243	.0975			16
103	.0292	.1268			17
102	.0342	.1609	.1545	.4554	18
101	.0392	.2001			19
100	.0439	.2441			20

TABLE 5

P = .75

	Singles		Batches of 3		
k	dJ= H(k)-H(k-1)	eJ= 1-H(k)	dJ	eJ	J
120	5E-11		.0003	.0003	0
119	7E-10				1
118	6E-9				2
117	3E-8		.0024	.0027	3
116	1.4E-7				4
115	5.5E-7				5
114	2.0E-6		.0100	.0127	6
113	5.0E-6				7
112	1.4E-5	0			8
111	3.0E-5	.0001	.0284	.0411	9
110	.0001	.0001			10
109	.0002	.0003			11
108	.0003	.0006	.0604	.1014	12
107	.0005	.0011			13
106	.0010	.0021			14
105	.0016	.0037	.1016	.2031	15
104	.0025	.0062			16
103	.0039	.0101			17
102	.0057	.0159	.1401	.3432	18
101	.0082	.0241			19
100	.0113	.0354			20
99	.0152	.0505	.1615	.5047	21
98	.0197	.0702			22
97	.0248	.0950			23
96	.0304	.1255	.1578		24
95	.0364	.1618			25

annually? If an annual series should end without clear resolution, as indeed it must occasionally according to the current plans what then? If there is not a clear cause of alarm, there is no need for alarm.

Consider a decision to limit the annual expenditure to 9 missiles, while extending the reporting period to cover 12 missiles (the current standard) if uncertainty had not been earlier resolved. In the worst case (all 12-missile series) reports would occur at 16-month intervals, or 8 reports in 11 years. Were the JCS to accept biennial reporting as an (occasional) substitute for annual reporting, this would be a technically simple solution.

Section 4 (Extract from Reference 17)

A Completely Bayesian Stopping Algorithm

[This is my suggestion for doing a complete Bayesian] decision theoretic analysis of the stopping problem. On the basis of the preliminary calculations described below, I estimate that this approach would reduce the number of units needed for testing by at least one per year over the savings which may be attained by using a sequential probability ratio test.

The approach requires the specification of a prior distribution and a loss structure. I suggest a possible form for these quantities below; but other choices would yield to similar analyses.

Let p denote the proportion of non-defective items in the population. Let h_1 denote a density on the unit interval, $0 < p < 1$; let h_0 denote the uniform density on the unit interval; and consider prior densities of the form

$$g(p) = w\, h_1(p) + (1-w)h_0(p), \tag{1}$$

where $0 < w < 1$ is a prior parameter. Here h_1 may be thought of as the posterior density which resulted from last year's tests, and w is the probability that p hasn't changed during the past year. If p has changed, which it may with probability 1-w, then it is assumed to be uniformly distributed over the interval $0 < p < 1$.

Suppose now that one may observe conditionally independent Bernoulli randon variables $X_1,\ldots,X_k$ with common success probability p, given p, and let

$$S_k = X_1+\ldots+X_k$$

denote the number of successes. Then the posterior distribution of p, given $X_1,\ldots,X_n$ is

$$g_k(p) = w\, h_1^k(p) + (1-w)h_0^k(p)$$

where $h_i^k(p) = h_i(p;k,S_k) \propto p^{S_k}(1-p)^{k-S_k}h_i(p)$

and $\int_0^1 h_i^k(p)\,dp=1$

Suppose now that a critical level p_0 is given with the following properties: if $p > p_0$, then the population contains enough good items; if $p < p_0$, then the population no longer contains enough good items and corrective action is desirable; and if p is much less than p_o, then corrective action is necessary. Suppose further that the purpose of each year's test is to decide whether $p < p_0$ or $p > p_0$; and define one unit of cost to be the cost of testing one item. Then the decision problem may be modelled as follows: the possible decisions are 1 to decided that $p < p_0$ and 2 to decide that $p > p_0$; if one decides that $p < p_0$ when, in fact, $p > p_0$, then one loses C_1 units; and if one decides that $p > p_0$, when, in fact, $p < p_0$, then one loses $C_2(p_0 - p)$ units. Here C_1 and C_2 are positive constants. C_1 represents the cost of inspecting the entire system; and the ratio C_2/C_1 is determined by the relative importance of the two kinds of errors.

These three elements, the prior distribution, the sampling distributions, and the loss structure, determine an optimal sampling plan, one which minimizes the sum of sampling costs and expected loss to due an incorrect decision. To describe it, first let m denote the maximum number of tests which could be conducted in any given year (e.g. m = 12). Next, let

$$L_1(k,s) = C_1 P(p > p_0 | S_k = s) + k$$

and $$L_2(k,s) = C_2 E\{\max(0, p_0 - p) | S_k = s\} + k$$

for $k = 0, \ldots, m$ and possible values of s. Thus L_1 and L_2 denote the conditional expected losses for the two decisions, given $X_1, \ldots, X_k$, plus the cost of observing $X_1, \ldots, X_k$. If $k = 0$, then $s = 0$ and the expectations are unconditional. If sampling is terminated after k tests, then it is optimal to make decision 1 if and only if $L_1(k,S_k) < L_2(k,S_k)$, in which the expected loss due to terminal decision is

$$L_0(k,S_k) = \min\{L_1(k,S_k), L_2(k,S_k)\}.$$

Let $p(k,s) = P(X_{k+1} = 1 \mid S_k = s)$

for $k = 1, \ldots, m-1$ and possible values of s; and define L by

$$L(m,s) = L_0(m,s)$$

and $$L(k,s) = \min\{L_0(k,s),$$

$$p(k,s)L(k+1,s+1) + (1-p(k,s))L(k+1,s)\} \tag{2}$$

for $k = 0, \ldots, m-1$ and possible values of s. Then the optimal sampling plan is to continue sampling as long as $L(k,S_k) < L_0(k,S_k)$, stopping at time

$$t = \min\{k>0: L_0(k,S_k) = L(k,S_k)\}.$$

Here $L(k,s)$ is the minimum expected loss plus sampling cost among all sampling plans which take at least k observations.

If h is a beta density, then it is possible to compute L_1 and L_2 as sums of products of p_0 and $(1-p_0)$ times ratios of factorials. I can supply the details, if you are interested. Using these explicit expressions, it is straightforward to compute L by the backward induction (2); and, once L and L_0 have been computed, it is simple to classify the possible outcomes (k,s) as stopping points, points for which $L_0(k,s) = L(k,s)$, or continuation points. Moreover, the stopping points divide themselves into lower stopping points for which $L_0(k,s) = L_1(k,s)$ and upper stopping points for which $L_0(k,s) = L_2(k,s)$. If the largest (smallest) lower (upper) stopping point is called a_k (resp. b_k), then

$$t = \min\{k>1: S_k < a_k \text{ or } S_k > b_k\}$$

and it is optimal to decide that $p < p_0$ if and only if $S_t < a_t$.

The several tables which accompany this letter describe the optimal sampling plan in a special case in which m = 12, h_1 is a beta density with parameters a = 6 and b = 2, w = 3/4, $p_0 = 3/4$, $C_1 = 60$, and $C_2 = 180$. Here the ratio $C_2/C_1 = 3$ equates the seriousness of deciding that $p < p_0$ when $p > p$ with that of deciding that $p > p_0$ when $p_0 - p = 1/3$; and the magnitudes of C_1 and C_2 were chosen to make it optimal to take up to about 12 observations. I believe that this is consistent with the power and sample size requirements discussed earlier. In a certain sense, these values of C_1 and C_2 are implicit in those requirements.

Table 1 lists the boundaries a_k and b_k of the optimal test. These boundaries are remarkably insensitive to a+b. I got nearly the same values when a = 9 and b = 3. Table 2 lists an ad hoc modification of the optimal boundaries which takes account of the economies of testing items in groups of three. Table 3 gives the posterior probability that $p > p_0$ for each possible outcome, using the adhoc boundaries. It clearly exhibits the following qualitative feature of the test: if the results of the first six tests this year are consistent with last year's results, then further testing is not optimal. Table 4 gives the frequentist properties of the adhoc test, the power function and expected sample size as a function of p. Observe that the maximum expected sample size is substantially smaller than that of the adhoc test; and recall the crucial role of the maximum in determining the number of items which must be purchased for testing.

TABLE 1: AN OPTIMAL BOUNDARY

Design Parameters: m=k, a=1, b=2, w=3/4, p=3/4, C1=60, C2=180

k	ak	bk
1	-	-
2	0	2
3	0	3
4	1	4
5	2	4
6	2	5
7	3	6
8	4	6
9	5	7
10	5	7
11	6	8
12	7	8

TABLE #2: A MODIFIED BOUNDARY

k	ak	bk
1	-	-
2	-	-
3	0	3
4	0	4
5	0	5
6	1	5
7	2	6
8	3	6
9	4	7
10	5	7
11	6	8
12	7	8

TABLE #3: POSSIBLE OUTCOMES WITH MODIFIED BOUNDARY

k	Sk	$P(p \geq po)$
3	0	.0251
6	1	.0084
6	2	.0507
7	3	.0813
8	4	.1211
9	5	.1634
11	6	.1185
12	7	.1546
12	8	.3111
10	7	.4543
8	6	.5183
6	5	.6517
3	3	.7450

TABLE #4: FREQUENTIST PROPERTIES

p	BETA	MEAN	VAR
.05	.9999	3.4575	1.281
.1	.999	3.8288	2.4702
.15	.9983	4.4161	3.5485
.2	.9903	4.8134	4.5345
.25	.9788	5.4154	5.43
.3	.9582	5.8102	6.2305
.35	.9244	6.3797	6.8348
.40	.8728	6.7887	7.559
.45	.8000	7.1384	8.1442

Comments by DW:

With this note Woodroofe completes the transition from Wald's classic treatment to a Bayesian approach. The use of a prior probability which is a mix of two hypotheses is in part an attempt to address the criticism that priors can become too sharply peaked, neglecting the potential staleness of old data. One might still ask whether there should be an upper limit to the value of k used in the prior.

The loss functions included in this section are representative, rather than my recommendation. The variable called po in the functions L1 and L2 could have different values in the two cases.

Chapter V

Other Stopping Criteria

A possible argument for small test sizes may arise after all missiles have been bought: any test reduces the potential tactical inventory. The decision criterion is unfortunately not unique. This chapter discusses a few examples.

Section 1. Utility as a Criterion

Let $\phi(p; s, f)dp$ be the posterior probability distribution of p, given s "equivalent" successes and f "equivalent" failures on which to base a prediction. Let U (N,p) be the "utility" of an inventory of N missiles of reliability p. The estimate of the utility of the inventory is then

$$U(N) = \int U(N,p)\,\phi(p; s, f)\,dp$$

Now perform a test: N goes to N-1; with probability p, s goes to s+1; and with probability 1-p, f goes to f+1.

After the test the utility is

$$U(N-1) = \int U(N-1, p)\left[p\phi(p; s+1, f) + (1-p)\phi(p; s, f+1)\right]dp.$$

The criterion is: Is U(N-1)>U(N)?

Examples of utility functions are:

Np (expected targets killed);

-Np(1-p) (uncertainty is reduced);

N-T/P (excess inventory, where T is size of critical target list);

$T\left[1-(1-p)^{N/T}\right]$ (expected damage);

$T\left[1-(1-ap)(1-p)^{b}\right]$ (b=largest integer in N/T; a=N/T-b is the fractional part; this reduces to Np for small N, goes to expected damage for large N).

Clearly there is a similarity between this method and that in Secion 4 of the previous chapter.

Section 2. Information as a Criterion

Another criterion would be the information the decision maker gains from the test about the posterior distribution of p. This would be applicable when no single utility function can be agreed on. An example is the Kullback-Leibler information measure on two probability density functions

F1 and F2 (Reference 18):

$$I(F_1, F_2) = \int F_1(p) \log \frac{F_1(p)}{F_2(p)} \, dp.$$

It can be applied to the current problem by defining F1 and F2 respectively as the posterior and prior density functions for p.

Shannon's information measure S(F1,F2) is the expectation value of I(F1,F2) over the observed values of success and failures.

To illustrate, we may identify F2 with expression 1.6 from Chapter I:

$$F_2(p) = p^{s_1-1}(1-p)^{f_1-1} / B(s_1, f_1)$$

and F1 with expression 1.8:

$$F_1(p) = p^{s_1+s_2-1}(1-p)^{f_1+f_2-1} / B(s_1+s_2, f_1+f_2)$$

so that log F1/F2 is

$$\log \frac{F_1(p)}{F_2(p)} = \log \left[\left\{ \frac{\Gamma(n_1+n_2)\Gamma(s_1)\Gamma(f_1)}{\Gamma(n_1)\Gamma(s_1+s_2)\Gamma(f_1+f_2)} \right\} p^{s_2}(1-p)^{f_2} \right]$$

$$= C + s_2 \log p + f_2 \log(1-p)$$

where C is the logarithm of the gamma-function combination in curly braces, all independent of p. Noting that

$$\int p^a \log p \, dp = \frac{\partial}{\partial a} \int p^a \, dp$$

and letting $\psi(z) = \frac{1}{\Gamma(z)} \frac{d\Gamma(z)}{dz}$, the logarithmic derivative of the gamma function, the expression for I(F1,F2) reduces to

$$I(F_1,F_2) = C - s_2\{\Psi(n_1+n_2) - \Psi(s_1+s_2)\} - f_2\{\Psi(n_1+n_2) - \Psi(f_1+f_2)\}.$$

Consider now the case where s2=n2=1 (a single successful trial). Then

$$I_S = \log\frac{n_1}{s_1} - \{\Psi(1+n_1) - \Psi(1+s_1)\}.$$

In the alternative case wher S2=0,n2=1 (a single unsuccessful trial)

$$I_F = \log\frac{n_1}{f_1} - \{\Psi(1+n_1) - \Psi(1+f_1)\}$$

and the Shannon information is $S = \dfrac{s_1 I_S + f_1 I_F}{n_1} \approx \dfrac{1}{2n_1} + \ldots$

As this never goes to zero (for finite n1), the cost of this information must be balanced against the use made of it.

I have not yet found a way to apply this criterion to the Pershing testing problem.

Chapter VI

Conclusion

I return now to the tasking from the Under Secretary of the Army, as given in the opening of this memorandum. The mathematical methods of sequential analysis proposed here for estimating reliability changes possess a rigor not found in the Army's current method, and make clear the risks in following their prescription. They provide a basis for reducing the size of an annual test and so reducing too the cost of a testing program. Indeed, they even challenge the need for an annual report, and suggest that the interval between reports can be enlarged (e.g., to two years) with no increase in risk to management. They do not, however, encompass a variety of other issues which are fundamentally operational in nature: firings to support training, alternate uses of inventory, system life. These must be the subject of further investigation.

Readers of this report may be disappointed that such very different approaches to the stopping problem have been presented in the foregoing chapters. I observe that such a seemingly simple problem has apparently not been hitherto subject to the scrutiny it deserves, and that it is comforting that two separate investigations have reached similar conclusions.

I see ultimately more promise in the methods proposed in Chapter IV, but would recommend that those of Chapters III and IV be applied to Pershing using the best available data so that a refined test program can be determined. In Chapter III is proposed the application, as yet unexplored, of Kalman filtering techniques to this problem. This research merits monitoring, if not support.

Appendix A

References

1. J. R. Ambrose (Under Secretary of the Army), "Operational Missile Testing". Memorandum to the Director of the Army Staff, 30 August 1982.

2. D. Willard, "Reliability Tracking by Sequential Sampling for Binomial Distributions", paper presented to 29th Conference on the Design of Experiments in Army Research and Development, October 1983.

3. Data Collection, Evaluation, and Reporting of Ballistic Missile Test Programs, JCS Memorandum SM-477-75 August 1975 (S).

4. Revised Guidelines for Use in Evaluating Strategic Ballistic Missile Operational Test Programs. IDA Study S-364/WSEG Report 92C, March 1975 (S).

5. JCS Guidelines for Missile Operational Testing and Reporting Memorandum SM-395-84, June 84 (S).

6. Revised Guidelines for Use in Evaluating Missile Operational Test Programs. Revision 92D to IDA Study S-364, February 1984 (S).

7. M. G. Kendall and A. Stuart, "The Advanced Theory of Statistics," Hafner, New York, 3rd Edition, 1969.

8. H. F. Martz and R. A. Waller "Bayesian Reliability Analysis", Wiley, New York, 1982.

9. H. Raiffa and R. Schlaifer, "Applied Statistical Decision Theory", Harvard University, 1961.

10. E. T. Jaynes, "Prior Probabilities", IEEE Transactions on Systems Science and Cybernetics, Vol. SS C-4, No. 3, Sept 1968.

11. A. Wald, "Sequential Analysis", Wiley, 1947. (Dover edition, 1973).

12. Andrew Hodges, "Alan Turing: The Enigma", Simon and Schuster, 1983.

13. M. W. Woodroofe, Unpublished Lecture Notes, October 1983.

14. M. W. Woodroofe, letter to R. Launer, August 2, 1983.

15. M. W. Woodroofe, letter to D. Willard, Novem 10, 1983.

16. M. W. Woodroofe, letter to D. Willard, December 14, 1983.

17. M. W. Woodroofe, Letter to D. Willard, January 17, 1984.

18. P. K. Goel and M. H. DeGroot "Information about Hyperparameters in Hierarchical Models." Journal of the American Statistical Association, Vol 76, p. 140, 1981.

Appendix B

Bibliography

Mann, Schafer, and Singpurwalla, "Methods for Statistical Analysis of Reliability and Life Data", Wiley, 1977.

C. P. Tsokos and I. N. Shimi (editors), "The Theory and Applications of Reliability", Academic Press, 1977.

J. J. Deeley, M. S. Tierney, and W. J. Zimmer, "On the Usefulness of the Maximum Entropy Principle in the Bayesian Estimation of Reliability, IEEE Transactions on Reliability, Vol. R-19, No. 3, August 1970.

Appendix C

Unclassified Extract from Reference 4:

Revised Guidelines for Use in Evaluating Stategic Ballistic Missile Operational Test Programs.

IDA Study S-364/WSEG Report 92 C, March 1975(S)

WSEG-92C, Cont.

D. ANALYSIS METHODOLOGY

(U) The various assumptions required in the [illegible] of the [illegible] analysis of the data should be specified. The mathematical methods and other data processing involved in deriving numerical performance estimates from the test data should be clearly defined for each performance measure [illegible] in the report. The [illegible] used in the calculations should be summarized to permit verification of the analytical approach.

E. SENSITIVITY ANALYSIS

(U) A sensitivity analysis should be conducted for each performance estimate to indicate whether the numerical results would change significantly if the treatment of test or data anomalies were changed.

F. CONFIDENCE STATEMENTS

(U) Two types of confidence statements should be provided for each performance factor:

(1) A statistical confidence bound based upon the *quantity* of data used in computing the factor.

(2) A qualitative assessment based upon the *quality* of data used in computing the factor.

The qualitative assessment should be based upon an appraisal of the validity and applicability of the test data as outlined in Part 1 of these guidelines.

(U) The statistical significance of differences in estimates of performance factors that is indicated by comparisons of the results of different sets of Operational Test data should be addressed and statistical confidence statements regarding these differences should be provided. The results of one method for comparing reliability samples is illustrated in Table 4.

Table 4(U). Statistical Significance of the Difference in Reliability Between Two Sets of Test Data

Data Set "A"	Data Set "B"		Difference in Reliability Between Data Sets "A" and "B"	Level of Significance of Difference in Reliability†
Reliability (Success Ratio)	No. of Tests	Reliability (Success Ratio)		
30/30 = 1.00	5	2/5 = .40	-.60	.99+
	10	4/10 = .40	-.60	.99+
	15	6/15 = .40	-.60	.99+
	5	3/5 = .60	-.40	.96
	10	6/10 = .60	-.40	.99+
	15	9/15 = .60	-.40	.99+
	5	4/5 = .80	-.20	.65
	10	8/10 = .80	-.20	.94
	15	12/15 = .80	-.20	.97
27/30 = .90	5	2/5 = .40	-.50	.97
	10	4/10 = .40	-.50	.99+
	15	6/15 = .40	-.50	.99+
	5	3/5 = .60	-.30	.86
	10	6/10 = .60	-.30	.95
	15	9/15 = .60	-.30	.98
	5	4/5 = .80	-.10	.54
	10	8/10 = .80	-.10	.63
	15	12/15 = .80	-.10	.69
	5	5/5 = 1.00	+.10	.38
	10	10/10 = 1.00	+.10	.59
	15	15/15 = 1.00	+.10	.71
24/30 = .80	5	1/5 = .20	-.60	.98
	10	2/10 = .20	-.60	.99+
	15	3/15 = .20	-.60	.99+
	5	2/5 = .40	-.40	.91
	10	4/10 = .40	-.40	.98
	15	6/15 = .40	-.40	.99
	5	3/5 = .60	-.20	.68
	10	6/10 = .60	-.20	.80
	15	9/15 = .60	-.20	.86
	5	5/5 = 1.00	+.20	.73
	10	10/10 = 1.00	+.20	.85
	15	15/15 = 1.00	+.20	.93

Table 4 (U). (Continued)

Data Set "A"*	Data Set "B"		Difference in Reliability Between Data Sets "A" and "B"	Level of Significance of Difference in Reliability†
Reliability (Success Ratio)	No. of Tests	Reliability (Success Ratio)		
21/30 = .70	5	1/5 = .20	-.50	.95
	10	2/10 = .20	-.50	.99
	15	3/15 = .20	-.50	.99+
	5	2/5 = .40	-.30	.79
	10	4/10 = .40	-.30	.91
	15	6/15 = .40	-.30	.99
	5	3/5 = .60	-.10	.49
	10	6/10 = .60	-.10	.59
	15	9/15 = .60	-.10	.74
	5	4/5 = .80	+.10	.45
	10	8/10 = .80	+.10	.57
	15	12/15 = .80	+.10	.63
	5	5/5 = 1.00	+.30	.80
	10	10/10 = 1.00	+.30	.95
	15	15/15 = 1.00	+.30	.99

*The number of tests in Data Set "A" is 30 for all cases shown.

†The values shown (F) are obtained by using Fisher's Exact Test:

$$P = 1 - \sum_{\nu = S_1}^{\nu_{max}} \binom{N_1}{\nu} \binom{N_2}{S_1 + S_2 - \nu} \bigg/ \binom{N_1 + N_2}{S_1 + S_2}$$

where $\binom{x}{y} = \frac{x!}{y!\,(x-y)!}$

$$\frac{S_1}{N_1} > \frac{S_2}{N_2}$$

$$\nu_{max} = \left.\begin{matrix} N_1 \\ S_1 + S_2 \end{matrix}\right\} \text{ whichever is smaller}$$

N_1 = number of tests in sample set 1

N_2 = number of tests in sample set 2

S_1 = number of successes in sample set 1

S_2 = number of successes in sample set 2

See A. Hald, *Statistical Theory With Engineering Applications*, John Wiley and Sons, Inc., 1960, p. 709.

Appendix D

HP-41 Programs

The HP-41 handheld calculator is slow but remarkably powerful. For example, a program listing for the standard Fast Fourier Transform (FFT) algorithm is no lengthier than that for a FORTRAN version and because of some quirks of the HP-41, the program is in some ways more efficient. With a 56-bit word, numerical accuracy is higher than in most personal computers, and so round-off problems are slower to arise.

Reported in this appendix are a set of programs written for this study. Their original purposes were to give or to verify solutions, but they have two additional values justifying their inclusion here: they demonstrate that the mathematics called upon is not intractible and can be packaged small, and they may be useful as is to others working the same or related problems.

The first group provde solutions to Equations 1.9 and 1.11 and thus can be considered a proper means of getting the answers wrongly sought via Fisher's Exact Test. The versions given are lengthy but are relatively robust to the accumulation of round-off errors. Included is the program PII, written to be a model for and to verify calculations of Singpurwalla and Launer.

The second group provide handy means of exploring Woodroofe's treatment of sequential analysis. ET provide solutions to Equations 1 and 2 of Chapter III, Sec 1. BND provides Wald's and Woodroofe's boundaries of the region of test continuation; and MW permits computation of a number of properties of a test plan defined by BND. LOP computes boundaries using the Bayesian method of Chapter III, Sec. 4.

Not included is a package of routines which manipulate truncated Taylor series and was used to compute the expansion of D(n,m) given in Eq 4.4. This is available from the author.

The memory requirements of an HP-41CV or CX are needed, and if it is not the CX version, then an Extended Functions module (XF) with its Expanded Memory. The occasional use of Synthetic Programming can be circumscribed, or if the programs are identical to those listed here, they should run on any version of the HP-41 with adequate memory and the XF module.

JCS+ · Implements Eq.1.9 and DA+ Eq.1.11.

They call for inputs and report the value of the integral as "CL=" for Confidence Level. The plus sign means there are no subtractions in the algorithm, hence less round-off error.

PII Implements Eqs.4-6 of Section III.3.

Entering at LBL A leads to an evaluation of α and at LBL B to evaluation of β . Lines 51-62 clear a block of registers, using program BC in a module called PPC ROM. This can be replaced by ordinary coding. If Flag 02 is set, then the summation sign in Eq.4 or 5 is ignored; only a single term is considered. Subroutines 1, 2, and 13 are the core of algorithm.

ET

Solves Eqs. 1 and 2 of Section IV.1.

$$\beta_0(p) = \sum_{k=c}^{N} \binom{N}{k} p^k (1-p)^{N-k} \qquad \text{and}$$

$$E(t_0) = \sum_{k=c}^{N} k \binom{k-1}{c-1} p^c (1-p)^{k-c} + \sum_{k=N-c+1}^{N} k \binom{k-1}{N-c} p^{k-N+c-1} (1-p)^{N-c+1}$$

$$= c p^c \sum_{0}^{N-c} \binom{k+c}{c} (1-p)^k + (N-c+1)(1-p)^{N-c+1} \sum_{0}^{c-1} \binom{k+N-c+1}{N-c+1} p^k.$$

Calls for N, c, and p (unadjusted values will be used as is).

Memory utilization keyed to that in MW: N, c, and p in same registers.

MW

Requires two files in Extended Memory named Am and Bm where m is a number provided in response to query "FILE#?" or is already stored in register 19. (Routine BND may have been used to create these files.)

Start program at line 1 or at LBL E; line one to provide/revise the value of N, the maximum number of tests. At E, provide "p" and "FILE#." If RAD-DEG selection set to RAD, program computes and reports G(k) as required by Section IV.3; if set to DEG, this is ignored.

Program reports β (p), E(t), and a (p) (which in effect interchanges meaning of "reliable" and "unreliable"). Sect IV.1.

LBL B produces output stating "bi/i = cumulative probability of sufficient failures to halt." Accumulates probability of exit passing clockwise around boundary. If there are several points on boundary at N=N max, then these are labeled F. Then program continues along "a" boundary.

LBL C does the same as LBL B but counterclockwise.

LOP

To meet the goals of Section IV.4. Computes the boundary conditions for continued testing, based on the loss functions L1 and L2 (which can have associated with them different criteria P1 and P2, as well as cost factors C1 and C2).

Program invites all necessary input insertion/revision/ verification, and then constructs a diagram cf the operating space. To conserve space this pattern is stored as packed binary data (a la flags). LBL J provides a visualization of this pattern, for display or printing (see figures below). This algorithm has also been run on a Commodore for verification.

Routines 6 and 7 support generation of loss functions L_1 and L_2 If others are chosen, these must be rewritten along with some of Routine 2 (lines 57-100).

BND

Develops the boundaries to be used in MW, by Wald's and Woodroofe's methods. Input called for: P0, P1, a, and b (later, m).

$0<P0<P1<1$. Level of test = a. Probability of Type II error = b

($P \geq P1$). Ho: $p \leq po$. (Section IV.2). M is number of tests.

Lines 1-85:	Wald's methods, a_n' and b_n' reported out.
86-156:	Woodroofe's modification.
157-END:	Subroutine E. Calls for a file number k; then stores Woodroofe's boundary numbers a_n and b_n in files AK and BK. If Flag 25 is clear to start, program halts if attempt is made to overwrite existing file. Set the Flag to permit overwriting.

JCS+

```
01♦LBL "JCS"
02 CF 29
03 "DEL="
04 SF 00
05 .
06 XEQ 00
07 "N1="
08  E
09 XEQ 00

10♦LBL B
11 "S1="
12 2
13 XEQ 00

14♦LBL C
15 "N2="
16 3
17 XEQ 00

18♦LBL D
19 "S2="
20 4
21 XEQ 00

22♦LBL 10
23 "REL DEG"
24 AVIEW
25 RCL 00
26 CHS
27  E
28 +
29 STO 11
30 RCL 04
31  E
32 +
33 RCL 03
34 -
35 STO 05
36 STO 06
37 LASTX
38  E
39 -
40 STO 08
41  E
42 -
43 RCL 02
44 +
45 STO 09
46 LASTX
47 CHS
48 RCL 01
49 +
50 STO 10

51♦LBL 01
52 RCL 06
53 STO 07

54♦LBL 02
55 RCL 06
56 RCL 07
57 -
58 LASTX
59  E
60 -
61 /
62 RCL 10
63 RCL 07
64 -
65 LASTX
66 RCL 09
67 +
68 /
69 *
70 RCL 00
71 /
72  E
73 X<> 13
74 *
75 ST+ 13
76 ISG 07
77 GTO 02
78 RCL 08
79 CHS
80 RCL 06
81 -
82 LASTX
83  E
84 -
85 /
86 RCL 00
87 RCL 11
88 /
89 *
90 RCL 13
91 X<> 12
92 *
93 ST+ 12
94 ISG 06
95 GTO 01

96♦LBL 03
97 RCL 11
98 RCL 08
99 Y↑X
100 ST* 12
101 RCL 02
102  E
103 -
104 RCL 03
105 +
106 LASTX
107 XEQ 04
108 ST* 12
109 RCL 01
110  E
111 -
112 RCL 03
113 +
114 LASTX
115 XEQ 04
116 ST/ 12
117 "CL="
118 FIX 4
119 ARCL 12
120 AVIEW
121 STOP
122 RTN

123♦LBL 00
124 FIX 0
125 FS?C 00
126 FIX 4
127 ARCL IND X
128 PROMPT
129 FS?C 22
130 STO IND Y
131 RTN

132♦LBL 04
133 CHS
134 X<>Y
135 SIGN
136 X<> L
137 ST+ Y

138♦LBL 05
139 X=Y?
140 GTO 06
141 ST* L
142 DSE X
143 GTO 05

144♦LBL 06
145 RDN
146 X<> L
147 RTN
148 .END.
```

DA+

```
01♦LBL "DA+"
02 CF 29
03 SF 00
04 "DEL="
05 .
06 XEQ 00

07♦LBL A
08 "H1="
09 E
10 XEQ 00

11♦LBL B
12 "S1="
13 2
14 XEQ 00

15♦LBL C
16 "H2="
17 3
18 XEQ 00

19♦LBL D
20 "S2="
21 4
22 XEQ 00

23♦LBL 10
24 "ABS DEG"
25 AVIEW
26 RCL 00
27 1/X
28 E
29 -
30 STO 09
31 RCL 01
32 RCL 02
33 -
34 STO 11
35 RCL 03
36 +
37 E
38 -
39 STO 05
40 RCL 04
41 RCL 03
42 -
43 E
44 +
45 STO 06
46 .
47 STO 15

48♦LBL 01
49 RCL 06
50 STO 07
51 RCL 03
52 X<>Y
53 +
54 STO 10
55 LASTX
56 E
57 +
58 RCL 05
59 +
60 STO 12

61♦LBL 02
62 RCL 12
63 RCL 07
64 -
65 STO 13
66 RCL 02
67 E
68 -
69 CHS
70 STO 08

71♦LBL 03
72 E
73 RCL 02
74 -
75 RCL 08
76 -
77 LASTX
78 E
79 -
80 /
81 RCL 10
82 RCL 08
83 -
84 LASTX
85 RCL 13
86 X<>Y
87 -
88 /
89 *
90 RCL 09
91 *
92 E
93 X<> 14
94 *
95 ST+ 14
96 ISG 08
97 GTO 03
98 RCL 06
99 RCL 07
100 -
101 LASTX
102 E
103 -
104 /
105 RCL 11
106 RCL 07
107 -
108 LASTX
109 RCL 12
110 X<>Y
111 -
112 /
113 *
114 RCL 09
115 *
116 RCL 14
117 X<> 15
118 *
119 ST+ 15
120 ISG 07
121 GTO 02
122 RCL 05
123 CHS
124 RCL 06
125 -
126 LASTX
127 E
128 -
129 /
130 RCL 09
131 /
132 RCL 15
133 X<> 16
134 *
135 ST+ 16
136 ISG 06
137 GTO 01

138♦LBL 04
139 RCL 00
140 RCL 02
141 E
142 -
143 Y↑X
144 ST* 16
145 RCL 00
146 CHS
147 E
148 +
149 RCL 05
150 Y↑X
151 ST* 16
152 RCL 01
153 E
154 -
155 RCL X
156 RCL 02
157 E
158 -
159 -
160 XEQ 05
161 ST* 16
162 RCL 05
163 RCL X
164 RCL 03
165 E
166 -
167 -
168 XEQ 05
169 ST/ 16
170 FIX 4
171 "CL="
172 ARCL 16
173 AVIEW
174 BEEP
175 STOP
176 RTN

177♦LBL 05
178 CHS
179 X<>Y
180 SIGN
181 X<> L
182 ST+ Y

183♦LBL 06
184 X=Y?
185 GTO 07
186 ST* L
187 DSE X
188 GTO 06

189♦LBL 07
190 RDN
191 X<> L
192 RTN

193♦LBL 00
194 FIX 0
195 FS?C 00
196 FIX 4
197 ARCL IND X
198 PROMPT
199 FS?C 22
200 STO IND Y
201 RTN
202 END
```

PII

Final Form

```
01 ♦LBL [illegible]
02 [illegible]
03 [illegible]
04 [illegible]
05 STO IND Y
06 RTN

07 LBL 09
08 RCL [illegible]
09 SF 01
10 X<> 03
11 [illegible] 15
12 X<> 03
13 RCL 16
14 RCL 17
15 -
16  E
17 -
18 STO 17
19 RDN
20 RTN

21♦LBL A
22 SF 00
23 GTO 10

24♦LBL B
25♦LBL "PII"
26 CF 00

27♦LBL 10
28 CF 01
29 FIX 2
30 CF 22
31 "DEL="
32 18
33 XEQ 00

34♦LBL C
35 "GAMMA="
36 3
37 XEQ 00
38 "DELTA="
39 15
40 XEQ 00

41♦LBL D
42 FIX 0
43 "N="
44 16
45 XEQ 00

46♦LBL E
47 FIX 0
48 "X="
49 17
50 XEQ 00
51 RCL 17
52 RCL 16
53  E
54 -
55  E3
56 /
57 FS? 02
58 X<>Y
59 STO 15
60 20.02
61 +
62 XROM 20.47
63 RCL 18
64 FS? 00
65 [illegible]
66 [illegible]
67 XEQ 09
68 ABS
69 STO 09

70♦LBL 07
71 RCL 19
72 INT
73 STO 17
74 RCL 09
75 X=0?
76 GTO 16
77 ENTER↑
78 CHS
79  E
80 +
81 STO 10
82 /
83 STO 08
84 - E
85 RCL 15
86 +
87 STO 14
88 LASTX
89 RCL 16
90 RCL 03
91 +
92 STO 04
93 +
94 STO 05
95 .
96 STO 13
97 RCL 16
98 STO 00
99 X=0?
100 GTO 20

101♦LBL 01
102 XEQ 21
103 RCL 16
104  E
105 +
106 RCL 14
107 RCL 04
108 RCL 00
109 ST- T
110 ST+ Z
111 ST- Y
112 *
113 /
114 *
115 RCL 09
116 /
117 ST* 13
118 DSE 00
119 GTO 01
120♦LBL [illegible]
121 XEQ 21
122 RCL 05
123 RCL 16
124 Y↑X
125 RCL 10
126 RCL 03
127 RCL 14
128 +
129 Y↑X
130 *
131 ST* 13
132 RCL 04
133  E
134 -
135 RCL 16
136 XEQ 04
137 ST* 13
138 RCL 05
139  E
140 -
141 RCL 16
142 XEQ 04
143 ST/ 13

144♦LBL 17
145 XEQ 19
146 RCL 12
147 ST/ 13
148 FIX 4
149 FC? 00
150 "b="
151 FS? 00
152 "1-a="
153  E
154 RCL 13
155 FC? 01
156 -
157 ARCL X
158 "⊢ X="
159 FIX 0
160 RCL 17
161 ARCL X
162 20
163 +
164 X<>Y
165 STO IND Y
166 AVIEW
167 ISG 19
168 GTO 07
169 FS? 01
170 XEQ 09
171 BEEP
172 STOP

173♦LBL J
174 20.02
175 RCL 19
176 FRC
177 +
178 XROM 20.07
179 RTN
```

"Block Clear"

"Block View"

```
181 SIGN
182 X<> L
183 X=0?
184 GTO 06
185 X<>Y

186♦LBL 05
187 ST* L
188 DSE X
189 ""
190 DSE Y
191 GTO 05

192♦LBL 06
193 RDN
194 X<> L
195 RTN

196♦LBL 21
197 RCL 16
198 RCL 00
199 -
200 STO 07
201 RCL 17
202 X>Y?
203 X<>Y
204 STO 01
205  E
206 ST+ 07
207 RCL 04
208 RCL 00
209 -
210 STO 06
211 RCL 03
212  E
213 STO 12
214 -
215 STO 02
216 X=0?
217 GTO 15

218♦LBL 02
219 RCL 05
220 RCL 03
221 RCL 06
222 RCL 02
223 ST- T
224 ST- Z
225 ST- Y
226 *
227 /
228 *
229 RCL 08
230 *
231 ST* 12
232  E
233 ST+ 12
234 DSE 02
235 GTO 02

236♦LBL 15
237 RCL 01
238 X=0?
239 GTO 14
240  E

241♦LBL 03
242 RCL 07
243 RCL 01
244 ST- Y
245 /
246 *
247 RCL 03
248 /
249  E
250 +
251 DSE 01
252 GTO 03
253 ST* 12

254♦LBL 14
255 RCL 12
256 ST+ 13
257 RTN

258♦LBL 16
259 CF 01
260 RCL 16
261  E
262 STO 13
263 +
264 STO 04
265 RCL 03
266  E
267 -
268 STO 05
269 RCL 16
270 RCL 15
271 +
272 STO 06
273 RCL 17
274 STO 00
275 X=0?
276 GTO 18

277♦LBL 13
278 RCL 04
279 RCL 07
280 RCL 06
281 RCL 00
282 ST- T
283 ST+ Z
284 ST- Y
285 *
286 /
287 *
288 ST* 13
289  E
290 ST+ 13
291 DSE 00
292 GTO 13

293♦LBL 18
294 RCL 14
295 RCL 16
296 +
297 LASTX
298 XEQ 04
299 ST* 13
300 RCL 05
301 RCL 06
302 +
304 XEQ 04
305 ST/ 13
306 GTO 17

307♦LBL 19
308 RCL 03
309  E
310 STO 11
311 -
312 STO 02
313 X=0?
314 GTO 92
315 RCL 09
316 X=0?
317 GTO 92
318 RCL 15
319  E
320 +

321♦LBL 91
322 ENTER↑
323 ENTER↑
324 RCL 08
325 *
326 RCL 02
327 /
328  E
329 X<> 11
330 *
331 ST+ 11
332 RDN
333 ISG X
334 ""
335 DSE 02
336 GTO 91

337♦LBL 92
338 RCL 09
339 CHS
340  E
341 +
342 RCL 14
343 RCL 03
344 +
345 Y↑X
346 RCL 11
347 *
348 STO 12
349 RTN

350♦LBL "PI"
351 STO 01
352 RCL 15
353 RCL 03
354 ST+ Y
355 X<>Y
356 /
357 RCL 09
358 -
359 STO [
360 ENTER↑
361 CHS
362  E
363 +
364 /
365  E

366♦LBL 03
367 RCL 16
368  E
369 +
370 RCL 01
371 ST- Y
372 /
373 R↑
374 *
375 *
376  E
377 +
378 DSE 01
379 GTO 03
380 RCL [
381 CHS
382  E
383 +
384 RCL 16
385 Y↑X
386 *
387 STOP
388 CHS
389  E
390 +
391 END
```

ET

```
01♦LBL "ET"
02 "N="
03 .
04 XEQ 00
05 "C="
06 E
07 XEQ 00
08 "P="
09 2
10 XEQ 00
11 SF 00
12 RCL 00
13 RCL 01
14 STO 04
15 -
16 STO 03
17 E
18 RCL 02
19 -
20 XEQ 01
21 RCL 02
22 RCL 01
23 Y↑X
24 *
25 RCL 01
26 *
27 STO 05
28 RCL 00
29 RCL 01
30 E
31 -
32 STO 03
33 -
34 STO 04
35 RCL 02
36 XEQ 01
37 E
38 RCL 02
39 -
40 RCL 00
41 RCL 01
42 -
43 E
44 +
45 Y↑X
46 LASTX
47 *
48 *
49 ST+ 05
50 "ET="
51 ARCL 05
52 AVIEW
53 CF 00
54 E
55 RCL 00
56 +
57 STO 04
58 LASTX
59 RCL 01
60 -
61 STO 03
62 E
63 RCL 02
64 -
65 LASTX
66 /
67 XEQ 01
68 RCL 02
69 RCL 00
70 Y↑X
71 *
72 STO 04
73 "B(P)="
74 4
75 XEQ 00
76 RTN

77♦LBL 00
78 CF 22
79 ARCL IND X
80 PROMPT
81 FS?C 22
82 STO IND Y
83 RTN

84♦LBL 01
85 ENTER↑
86 ENTER↑
87 ENTER↑
88 E

89♦LBL 02
90 *
91 RCL 04
92 RCL 03
93 FS? 00
94 +
95 FC? 00
96 -
97 LASTX
98 /
99 *
100 E
101 +
102 DSE 03
103 GTO 02
104 END
```

Output of MW

LBL B

```
n3/1=0.0000
3/2=0.0000
3/3=0.0034
4/4=0.0034
4/5=0.0047
4/6=0.0080
4/7=0.0139
4/8=0.0228
4/9=0.0352
4/10=0.0509
4/11=0.0695
4/12=0.0899
a12=0.2056
11=0.3776
10=0.3776
9=0.6229
8=0.6229
7=0.6229
6=1.0000
5=1.0000
4=1.0000
3=1.0000
2=1.0000
1=1.0000
```

LBL C

```
a-1/1=0.0000
-1/2=0.0000
-1/3=0.0000
-1/4=0.0000
-1/5=0.0000
0/6=0.3771
0/7=0.3771
0/8=0.3771
1/9=0.6224
1/10=0.6224
2/11=0.7944
3/12=0.9101
b12=0.9305
11=0.9491
10=0.9648
9=0.9772
8=0.9861
7=0.9920
6=0.9953
5=0.9966
4=0.9966
3=1.0000
2=1.0000
1=1.0000
```

BND

Constructs and displays boundaries for input to MW, MX. Those labeled a'_n & b'_n are computed using Wald's equations; a_n & b_n using Woodroofe's mods. (mano 10 Mar 83)

Subroutine E stores a_n & b_n

Wald derives a'_n & b'_n, $1 \le n \le M$. Woodroofe then picks a_M & b_M ∋ $a'_M \le a_M = b_M - 1$, $b_M \le b'_M$, $a'_M + b'_M \simeq a_M + b_M$

$a_n = \max\{a'_n, a_M - M + n\}$

$b_n = \min\{b'_n, b_M\}$ $(n < M)$.

$0 < P_0 < P_1 < 1$

a = level of test.

b = prob type II error (P ≥ P₁)

H_0: P ≤ P_0.

```
01•LBL "BND"
02 "P0"
03 .
04 FIX 4
05 XEQ 00
06 "P1"
07 E
08 XEQ 00
09 "a"
10 2
11 XEQ 00
12 "b"
13 3
14 XEQ 00
15 E
16 RCL 02
17 -
18 RCL 03
19 /            A
20 LN
21 STO 04
22 E
23 RCL 03
24 -
25 RCL 02
26 /            B
27 LN
28 STO 06
29 E
30 RCL 00
31 -
32 E
33 RCL 01
34 -
35 /
36 LN
37 STO 05       Δ0
38 RCL 01
39 RCL 00
40 /
41 LN
42 +            Δ1
43 ST/ 04 (log A)/Δ1
44 ST/ 05 Δ0/Δ1
45 ST/ 06 (log B)/Δ1
46 FIX 0
47 "M"
48 7
49 XEQ 00
```

```
50 RCL 07
51 E3
52 /
53 E
54 +
55 STO 08 (1.00M)
56 STO 11
57 RCL 04
58 CHS
59 STO 09 ln 1/A
60 "a'"
61 ASTO 10
62 20
63 STO 12 Ra'
64 RCL 05
65 ENTER↑
66 ENTER↑
67 ENTER↑
68 XEQ 01
69 RCL 11
70 STO 08
71 RCL 06
72 E
73 +
74 STO 09 (ln B)+1
75 "b'"
76 ASTO 10
77 20
78 RCL 07
79 +
80 STO 12 Rb'
81 RCL 05
82 ENTER↑
83 ENTER↑
84 ENTER↑
85 XEQ 01      WALD
86 "a"         Woodroofe
87 ASTO 10
88 RCL 07
89 STO 03
90 ENTER↑
91 ENTER↑
92 ENTER↑
93 19
94 +
95 STO 13 Ra'M
96 +
97 +
98 STO 12 RaM
99 STO 14
100 "aM=?"
101 PROMPT
102 STO IND 12
103 CF 00
```

```
104•LBL 02
105 RCL IND 13
106 X<Y?
107 SF 00
108 X<Y?
109 X<>Y
110 XEQ 04
111 DSE X
112 --
113 DSE 03
114 GTO 02
115 "b"
116 ASTO 10
117 RCL 07
118 STO 03
119 ST+ X
120 ENTER↑
121 ENTER↑
122 19
123 +
124 STO 13
125 +
126 STO 12
127 STO 15
128 "bM=?"
129 PROMPT
130 STO IND 12
131 CF 00

132•LBL 03
133 RCL IND 13
134 X>Y?
135 SF 00
136 X>Y?
137 X<>Y
138 XEQ 04
139 DSE 03
140 GTO 03
141 RTN

142•LBL 04
143 STO IND 12
144 CLA
145 ARCL 10
146 ARCL 03
147 "⊢="
148 ARCL X
149 AVIEW
150 FC?C 00
151 RDN
152 DSE 13
153 --
154 DSE 12
155 --
156 RTN
```

To store a_n & b_n. Calls for m in A_m, B_m ID's of files.

```
157•LBL E
158 "FL#?"
159 PROMPT
160 STO 11
161 RCL 14
162 "A"
163 XEQ 03
164 RCL 15
165 "B"

166•LBL 05
167 RCL 07
168 ARCL 11
169 CRFLD
170 CF 25
171 .
172 SEEKPTR
173 X<> Z
174 1.001
175 *
176 X<>Y
177 -
178 E
179 +
180 SAVERX
181 RTN
```

Reports "DUP FL" if duplicate exists. SF25 to overwrite. R/S or restart E, using new #.

```
182•LBL 01
183 RCL 09
184 +
185 STO 09
186 RCL X
187 E
188 MOD
189 -
190 STO IND 12
191 CLA
192 ARCL 10
193 ARCL 08
194 "⊢="
195 ARCL X
196 RDN
197 AVIEW
198 ISG 12
199 --
200 ISG 08
201 GTO 01
202 RTN
```

Replaces INT to max[X] whether + or −.

```
203•LBL 00
204 CF 22
205 "⊢="
206 ARCL IND ...
207 PROMPT
208 FS? 22
209 STO IND ...
210 END
```

Final Version

RAD (vs DEG) used as flag to determine whether to execute lines 135-157.

MW

```
01•LBL "MM"
02 CF 29
03 "M="
04 .
05 XEQ 00
06 RCL 00
07 E
08 -
09 STO 15

10•LBL E
11 CF 21
12 FIX 4
13 "P="
14 2
15 XEQ 00
16 E
17 RCL 02
18 -
19 STO 03
20 "FILE#="
21 19
22 FIX 0
23 XEQ 00
24 22
25 STO 04
26 E
27 RCL 00
28 +
29 +
30 STO 05
31 LASTX
32 +
33 STO 06
34 PSIZE
35 "A"
36 RCL 04
37 RCL 05
38 XEQ 20
39 "B"
40 RCL 05
41 RCL 06
42 XEQ 20
43 RCL IND X
44 ISG Y
45•LBL 01
46 RCL IND Y
47 X>Y?
48 X<>Y
49 RDN
50 ISG Y
51 GTO 01
52 STO 01
53 E
54 RCL 06
55 +
56 +
57 PSIZE
58 E
59 STO IND L
60 RCL 15
61 E3
62 /
63 STO 20
64 CLX
65 STO 10
66 STO 11
67 STO 12

68•LBL 02
69 RCL 20
70 INT
71 ENTER↑
72 ENTER↑
73 ENTER↑
74 RCL 04
75 +
76 STO 07
77 E
78 +
79 STO 17
80 RDN
81 RCL 05
82 +
83 STO 08
84 E
85 +
86 STO 18
87 RDN
88 X=0?
89 GTO 04
90 RCL 01
91 E
92 -
93 X>Y?
94 X<>Y
95 STO 21
96 E3
97 /
98 STO 14
99 CLX
100 X<>F

101•LBL 03
102 RCL 21
103 RCL 14
104 INT
105 -
106 STO 13
107 RCL 06
108 +
109 E
110 +
111 STO 09
112 RCL IND 07
113 RCL 13
114 X<=Y?
115 SF 05
116 RCL IND 08
117 X>Y?
118 SF 06
119 RCL 09
120 RCL 03
121 FS? 05
122 CLX
123 ST* IND Y
124 DSE Y
125 RCL IND Y
126 RCL 02
127 FC? 06
128 CLX
129 *
130 ST+ IND 09
131 ISG 14
132 GTO 03
133 FC? 43
134 GTO 04
135 RCL 09
136 RCL 06
137 RCL 01
138 +
139 E3
140 /
141 +
142 .

143•LBL 14
144 RCL IND Y
145 +
146 ISG Y
147 GTO 14
148 CHS
149 E
150 +
151 "G"
152 FIX 0
153 ARCL 20
154 "F="
155 FIX 4
156 ARCL X
157 AVIEW

158•LBL 04
159 RCL 06
160 RCL IND 18
161 +
162 STO 18
163 RCL IND X
164 RCL 02
165 *
166 STO 16
167 ST+ 10
168 STO IND 08
169 RCL 06
170 RCL IND 17
171 +
172 E
173 +
174 STO 17
175 RCL IND X
176 RCL [illegible]
177 *
178 ST+ 11
179 STO IND 07
180 RCL 15
181 +
182 RCL 20
183 INT
184 E
185 +
186 *
187 ST+ 12
188 ISG 20
189 GTO 02
190 RCL 17
191 RCL 18
192 X<=Y?
193 GTO 06
194 RCL 03
195 ST* IND Y
196 RDN
197 RCL 02
198 ST* IND Z
199 RDN
200 E3
201 /
202 +
203 STO 19

204•LBL 05
205 RCL IND 19
206 RCL 00
207 *
208 ST+ 12
209 ISG 19
210 GTO 05

211•LBL 06
212 SF 21
213 BEEP
214 FIX 4
215 "D(P)="
216 10
217 [illegible]
218 [illegible]
```

LBL C is LBL B run counterclockwise
around the border.

```
219 11
220 XEQ 00
221 "E(T)="
222 12

223•LBL 00
224 CF 22
225 ARCL IND X
226 PROMPT
227 FS?C 22
228 STO IND Y
229 RTN

230•LBL 20
231 ARCL 19
232 .
233 SEEKPTA
234 SIGN
235 ST+ Z
236 -
237 E3
238 /
239 +
240 GETRX
241 RTN

242•LBL B
243 "B"
244 FIX 0
245 ARCL 19
246 CLX
247 SEEKPTA
248 X<>F
249 RCL 05
250 RCL 00
251 E3
252 /
253 E
254 +
255 .
256 "b"
257 XEQ 09
258 SF 00
259 SF 01
260 RCL 17
261 CHS
262 RCL 18
263 +
264 X<=0?
265 GTO 07
266 LASTX
267 X<>Y
268 E
269 +
270 RT
271 "F"
272 XEQ 09
273 X<>Y

274•LBL 07
275 X<>Y
276 RCL 05
277 2
278 -
279 RCL 00
280 RCL Z
281 "a"
282 XEQ 09
283 RTN

284•LBL C
285 "H"
286 FIX 0
287 ARCL 19
288 CLX
289 SEEKPTA
290 X<>F
291 RCL 04
292 RCL 00
293 E3
294 /
295 E
296 +
297 .
298 "a"
299 XEQ 09
300 SF 01
301 RCL 18
302 RCL 17
303 -
304 X<=0?
305 GTO 08
306 LASTX
307 X<>Y
308 E
309 +
310 E3
311 /
312 E
313 +
314 RT
315 "F"
316 XEQ 09
317 X<>Y

318•LBL 08
319 X<>Y
320 RCL 06
321 2
322 -
323 RCL 00
324 RCL Z
325 "b"
326 SF 00

327•LBL 09
328 FIX 0
329 RCL IND Z
330 +
331 FS? 01
332 GTO 10
333 GETX
334 ARCL X
335 RDN
336 "F/"

337•LBL 10
338 ARCL Y
339 "F="
340 FIX 4
341 ARCL X
342 AVIEW
343 CLA
344 FS? 09
345 GTO 11
346 ISG Z
347 --
348 ISG Y
349 GTO 09
350 RTN

351•LBL 11
352 DSE Z
353 --
354 DSE Y
355 GTO 09
356 .END.
```

ho d Vessen, Sering on [illegible]

cells

LOP

```
01◆LBL "LOP"
02 "M"
03 CLX
04 XEQ 00
05 "a"
06 3
07 XEQ 00
08 "b"
09 4
10 XEQ 00
11 "P"
12 5
13 XEQ 00
14 "W"
15 6
16 XEQ 00
17 "C1"
18 17
19 XEQ 00
20 "C2"
21 18

22◆LBL 00
23 CF 22
24 "F="
25 ARCL IND Y
26 AVIEW
27 FS?C 22
28 STO IND Y
29 RTN
```

Endyfrequency

```
30◆LBL E
31 20
32 STO 05
33 PSIZE
34 RCL 00
35 ENTER↑
36 ISG X
37 ""
38 ENTER↑
39 ISG X
40 ""
41 *
42 2
43 /
44 R↑
45 +
46 STO 11
47 +
48 PSIZE        Set-up
49 RCL 00
50 E
51 ST- 11
52 +
53 STO 01
```

$k_0 = m$

(Because only DSE while [illegible])

Compute L_1, L_2 & $\min(L_1, L_2)$

```
54◆LBL 01
55 RCL 01
56 STO 02

57◆LBL 02
58 RCL 03
59 STO 07
60 RCL 04
61 STO 08
62 XEQ 06
63 RCL 06
64 *
65 STO 13
66 XEQ 07
67 RCL 06
68 *
69 STO 14
70 E
71 STO 07
72 STO 08
73 XEQ 06
74 E
75 RCL 05
76 -
77 *
78 RCL 13
79 +
80 RCL 17
81 *
82 RCL 01
83 +
84 E
85 -
86 STO 13
87 XEQ 07
88 E
89 RCL 06
90 -
91 *
92 RCL 14
93 +
94 RCL 18
95 *
96 RCL 01
97 +
98 E
99 -              L2
100 RCL 13        L1
101 X>Y?
102 X<>Y
103 STO 14        Min(L1,L2)
104 RCL 00
105 RCL 02
106 2
107 -
```

```
108 LASTX
109 /
110 -
111 RCL 02
112 E
113 -
114 *
115 RCL 01
116 +
117 E
118 -
119 RCL 09
120 +
121 R↑
122 STO IND Y     Store min(L1,L2) (retained in stack till once)
123 DSE 02
124 GTO 02
125 DSE 01
126 GTO 01
127 RCL 00
128 STO 01
129 SF 21
130 RCLFLAG
131 .             Clear All Flags
132 STO d
133 X<>Y
134 21.43         Restore only 21-43
135 STOFLAG
136 RCLFLAG       Save 21-43, 0-20 Clear
137 STO 16

138◆LBL 03
139 RCL 01
140 STO 02

141◆LBL 04
142 RCL 02
143 RCL 03
144 E
145 -
146 +
147 LASTX
148 RCL 01
149 +
150 RCL 04
151 +
152 /
153 RCL 06
154 *
155 E
156 LASTX
157 -
158 RCL 02
159 *
```

```
160 RCL 01
161 E
162 +
163 /
164 +
165 STO 15       p(k,s)
166 -!
167 RCL 02
168 +
169 RCL X
170 LASTX
171 *
172 -2
173 /
174 ST+ Y
175 X<>Y
176 RCL 02
177 RCL 00
178 *
179 ST+ Z
180 LASTX
181 -
182 +
183 RCL 01
184 RCL 09
185 +
186 ST+ Z
187 +
188 STO 10
189 RCL IND Y
190 RCL IND Y
191 E
192 ST- 10
193 RCL 15
194 ST* T
195 -
196 *
197 +
198 RCL IND 10
199 X<>Y
200 X<Y?
201 STO IND 10
202 X<Y?
203 SF IND 02
204 DSE 02
205 GTO 04
206 RCL 11
207 RCL 01
208 +
209 RCLFLAG
210 STO IND Y
211 RCL 16
212 STOFLAG
213 DSE 01
214 GTO [illegible]
215 BEEP
216 STOP
```

Reports Continuation Cells for Viewing & Printing

```
217♦LBL J
218 RCL 00
219 RCL 11
220 +
221 LASTX
222 X<>Y
223  E3
224 /
225 +
226  E
227 +
228 STO 01

229♦LBL 14
230 RCL 00
231  E3
232 /
233  E
234 +
235 STO 02
236 RCL IND 01
237 STOFLAG
238 CLA
239 RCL b
240 FS? IND 03
241 "├*"
242 FC? IND 03
243 "├!"
244 ISG 02
245 STO b
246 SF 12
247 AVIEW
248 ISG 01
249 GTO 14
250 CF 12
251 CLX
252 RTN
```

127,42

127,33

Can put 127,32 here instead of SF12

Better: RCL16 STOFLAG INSTEAD

Compute part of b_1

```
253♦LBL 06
254 RCL 01
255 RCL 07
256 +
257 RCL 06
258 +
259 2
260 -
261 STO ]
262  E
263 RCL 05
264 ST- Y
265 [illegible]
266 [illegible]
267 STO [illegible]
268 RCL 02
269 RCL 07
270 +
271 2
272 -
273 STO [
274 CF 00
275 X=0?
276 SF 00
277  E
278 FS?C 00
279 GTO 11
280 ENTER↑
281 ENTER↑
282 ENTER↑

283♦LBL 10
284 RCL \
285 *
286 RCL ]
287  E
288 +
289 RCL [
290 ST- Y
291 /
292 *
293 +
294 DSE [
295 GTO 10

296♦LBL 11
297  E
298 RCL 05
299 -
300 RCL ]
301 Y↑X
302 *
303 RTN
```

Compute part of b_2

```
304♦LBL 07
305  E
306 RCL 05
307 ST- Y
308 /
309 STO \
310 RCL 01
311 RCL 07
312 +
313 RCL 03
314 +
315  E
316 -
317 STO [illegible]
318 STO [illegible]
319 RCL 02
320 RCL 07
321 +
322  E
323 ST+ ]
324 RDN
325 ST- [
326 STO ↑
327 CF 00
328 X<>Y
329 X<=Y?
330 SF 00
331 R↑
332 FS?C 00
333 GTO 13
334 ENTER↑
335 ENTER↑
336 ENTER↑

337♦LBL 12
338 RCL \
339 *
340 RCL [
341 /
342 +
343 RCL ]
344 RCL [
345 -
346 *
347 RCL ↑
348 -
349 +
350 DSE [
351 GTO 12

352♦LBL 13
353 RCL 05
354 RCL ]
355  E
356 -
357 Y↑X
358 LASTX
359 /
360 *
361 END
```

503 bytes

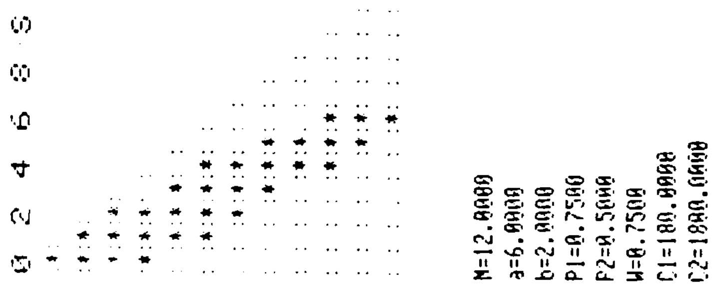
Sample output of Subroutine J

```
M=12.0000
a=6.0000
b=2.0000
P=0.7500
W=0.5000
C1=60.0000
C2=180.0000

* ! ! ! ! ! ! ! ! ! ! !
* * ! ! ! ! ! ! ! ! ! !
! * * ! ! ! ! ! ! ! ! !
! ! * ! ! ! ! ! ! ! ! !
! ! * * ! ! ! ! ! ! ! !
! ! ! * * ! ! ! ! ! ! !
! ! ! ! * ! ! ! ! ! ! !
! ! ! ! * * ! ! ! ! ! !
! ! ! ! ! * * ! ! ! ! !
! ! ! ! ! ! * ! ! ! ! !
! ! ! ! ! ! ! * ! ! ! !
! ! ! ! ! ! ! ! ! ! !
```

```
M=12.0000
a=6.0000
b=2.0000
P=0.5000
W=0.9000
C1=60.0000
C2=180.0000

! ! ! ! ! ! ! ! ! ! ! !
! ! ! ! ! ! ! ! ! ! ! !
* ! ! ! ! ! ! ! ! ! ! !
* ! ! ! ! ! ! ! ! ! ! !
* ! ! ! ! ! ! ! ! ! ! !
* ! ! ! ! ! ! ! ! ! ! !
* * ! ! ! ! ! ! ! ! ! !
* * ! ! ! ! ! ! ! ! ! !
! * ! ! ! ! ! ! ! ! ! !
! * * ! ! ! ! ! ! ! ! !
! ! * ! ! ! ! ! ! ! ! !
! ! ! ! ! ! ! ! ! ! ! !
```

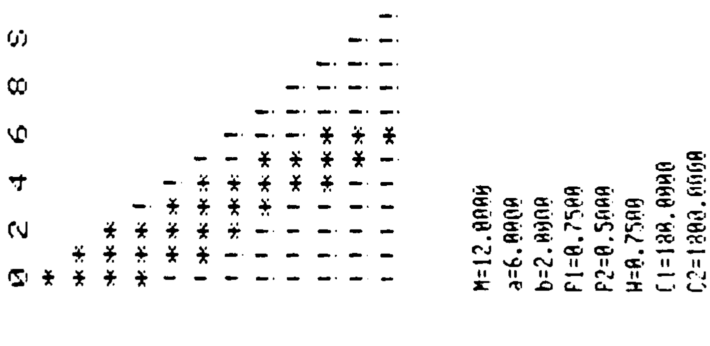
```
M=12.0000
a=6.0000
b=2.0000
P=0.7500
W=0.7500
C1=400.0000
C2=600.0000
                          RUN

                          RUN
* ! ! ! ! ! ! ! ! ! ! !
* * ! ! ! ! ! ! ! ! ! !
* * * ! ! ! ! ! ! ! ! !
* * * * ! ! ! ! ! ! ! !
* * * * ! ! ! ! ! ! ! !
! * * * * ! ! ! ! ! ! !
! ! * * * ! ! ! ! ! ! !
! ! ! * * * ! ! ! ! ! !
! ! ! ! * * * ! ! ! ! !
! ! ! ! ! * * ! ! ! ! !
! ! ! ! ! ! * * ! ! ! !
! ! ! ! ! ! ! * ! ! ! !
```

```
M=12.0000
a=6.0000
b=2.0000
P=0.7500
W=0.7500
C1=60.0000
C2=180.0000

* ! ! ! ! ! ! ! ! ! ! !
* * ! ! ! ! ! ! ! ! ! !
! * * ! ! ! ! ! ! ! ! !
! * * ! ! ! ! ! ! ! ! !
! ! * * ! ! ! ! ! ! ! !
! ! ! * ! ! ! ! ! ! ! !
! ! ! ! * ! ! ! ! ! ! !
! ! ! ! * * ! ! ! ! ! !
! ! ! ! ! * ! ! ! ! ! !
! ! ! ! ! ! * ! ! ! ! !
! ! ! ! ! ! ! * ! ! ! !
! ! ! ! ! ! ! * ! ! ! !
```



Other versions of output, making clearer the upper boundary.
Last program version will not print the last cell (i.e, won't indicate S=12, K=12) but an unlikely case.

Note that P∅ is now replaced by P1 and P2



Appendix E

The appendices to Chapter III.

APPENDIX A

An Algorithm, A Computer Code, and A User's Guide, for a Bayesian Binomial Hypothesis Testing Procedure

A.1. INTRODUCTION

In the Bayesian binomial hypothesis testing procedure, we need to find the pair (n_t, x_t^*) such that [see Equations (4) and (5)]:

$$\int_0^1 \sum_{j=0}^{x_t^*} \binom{n_t}{j} p_t^j (1 - p_t)^{n_t - j} g(p_t)dp_t \leqslant \alpha$$

and

$$\int_\Delta^1 \sum_{j=0}^{x_t^*} \binom{n_t}{j} (p_t - \Delta)^j (1 - p_t + \Delta)^{n_t - j} g(p_t)dp_t \geqslant 1 - \beta ,$$

where

$$g(p_t) = \frac{\Gamma(\gamma+\delta)}{\Gamma(\gamma)\Gamma(\delta)} p_t^{\gamma-1} (1 - p_t)^{\delta-1} .$$

The above inequalities can be rewritten as:

$$g_1(x_t^*, n_t) = \frac{\Gamma(\gamma+\delta)}{\Gamma(\gamma)\Gamma(\delta)} \sum_{j=0}^{x_t^*} \binom{n_t}{j} \frac{\Gamma(j+\gamma)\Gamma(n_t-j+\delta)}{\Gamma(n_t+\gamma+\delta)} , \tag{8A}$$

$$g_2(x_t^*, n_t) = \frac{\Gamma(\gamma+\delta)}{\Gamma(\gamma)\Gamma(\delta)} \Delta^{n_t} \sum_{j=0}^{x_t^*} \binom{n_t}{j} \left[\sum_{\ell=0}^{j} \binom{j}{\ell} \Delta^{-\ell} (-1)^{j-\ell} \cdot \left\{ \sum_{m=0}^{n_t-j} \binom{n_t-j}{m} \Delta^{-m} B(\Delta,1;\ \ell+\delta,\ m+\delta) \right\} \right] \geqslant 1 - \beta , \tag{9A}$$

where

$$B(\Delta,1;\, r,s) \;=\; \int_{\Delta}^{1} p_t^{r-1}\,(1 - p_t)^{s-1}\,dp_t \;.$$

A computer code designed to obtain the smallest values of n_t, x_t^* subject to the two inequalities (8A) and (9A), based on an enumeration procedure discussed next, is obtained.

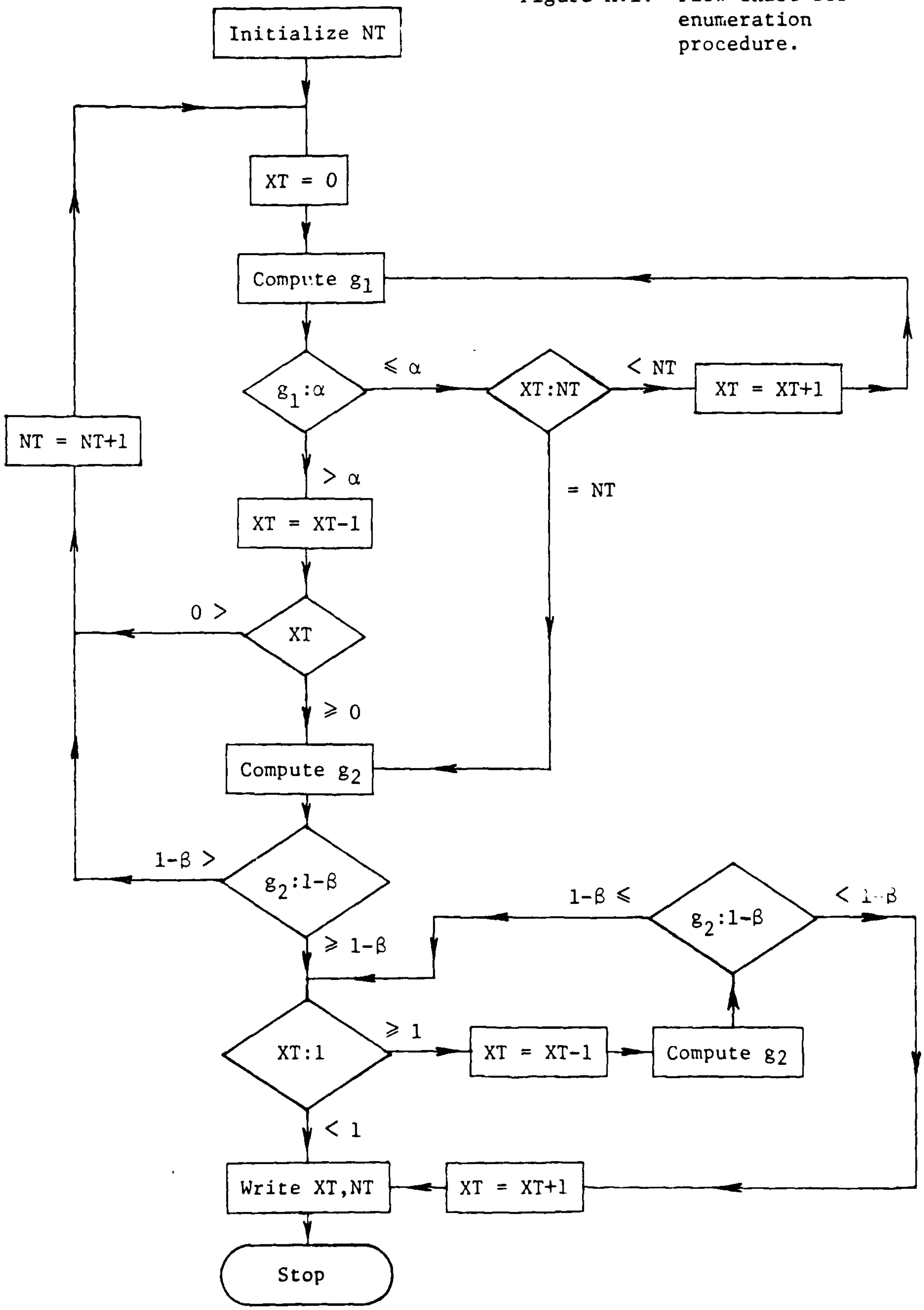
A.2 DESCRIPTION OF THE ENUMERATION PROCEDURE

The enumeration procedure exploits the fact that both $g_1(x_t,n_t)$ and $g_2(x_t,n_t)$ are increasing functions of x_t if n_t is fixed. The procedure starts with some initial value of n_t, say n_t^0, and finds the largest x_t such that $g_1(x_t,n_t^0) \leqslant \alpha$. Once such an x_t, say x_t^0, is found, it is guaranteed that the first inequality will be satisfied for values of x_t smaller than x_t^0. The procedure then tries to find an x_t smaller than x_t^0 such that $g_2(x_t,n_t) \geqslant 1 - \beta$. If such an x_t does not exist, the value of n_t is increased by one and the procedure starts all over again. As n_t increases, the procedure finds the smallest values of n_t and x_t satisfying both inequalities. The flow chart for this enumeration procedure is presented in Figure A.1.

A.3 THE COMPUTER CODE

The program requires certain JCL cards and a user input of some parameters.

Figure A.1. Flow chart for enumeration procedure.



A.3.1 Input Specifications

The cards should be arranged in accordance with Figure A.2; each card will be explained individually.

Job Card and JCL Cards: The standard job card is used and so are the following JCL cards:

```
//bEXECbFORG2
//FORT.SYSINbDD
//GO.SYSLIBbDD
//bbbbbDDbbbbDSN=GWU.IMSL.V9.DLOAD,DISP=SHR
//GO.SYSINbbbbbDDbbbbb*
```

where the character "b" indicates a blank space. The first two JCL cards immediately follow the job card. The remaining JCL cards are placed after the program and just before the input information card. The fourth JCL card is needed to use the IMSL subroutines on an IBM machine.

Input Information Card--DEL, SGM, SDEL, ALF, BETA, NT: This card contains sorted input information, DEL, SGM, and SDEL, which are the parameters Δ , γ , and δ in Equations (8A) and (9A); ALF and BETA are the right-hand side parameters α and β in these inequalities. These parameters are specified in format F10.5. The input NT is the initial value of n_t selected, and is in I4 format. Usually, this value is one.

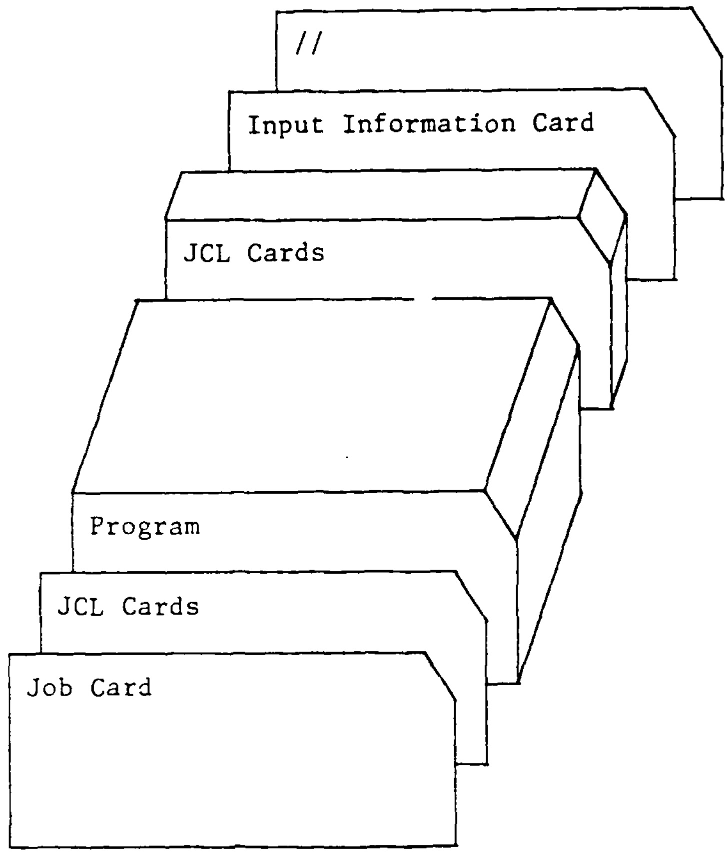


Figure A.2. Card deck structure.

A.3.2 Interpretation of Output

The program uses an iterative scheme and evaluates $g_1(x_t,n_t)$ and $g_2(x_t,n_t)$ for different values of x_t adn n_t. On the output, the values of $g_1(x_t,n_t)$ and $g_2(x_t,n_t)$ are printed as

FIRST CONST =

SECOND CONST =

for different values of x_t and n_t.

The solution of the problem, that is, the smallest values of x_t and n_t satisfying the inequalities (8A) and (9A), are printed in the

last line of the output as

X =   N =

Sample output is presented in Table A.1.

The smallest values of x_t and n_t satisfying the inequalities (8A) and (9A) are X = 10 and N = 15 . In this example, the values of the parameters are $\Delta = 0.25$, $\gamma = 106$, $\delta = 19$, $\alpha = 0.10$, and $\beta = 0.25$. The initial value of n_t is one.

The listing of the program is given in Appendix B.

TABLE A.1

Sample Output

FIRST CONST=	0.00000	XT=	0.0	NT=	12
FIRST CONST=	0.00000	XT=	1.0	NT=	12
FIRST CONST=	0.00000	XT=	2.0	NT=	12
FIRST CONST=	0.00002	XT=	3.0	NT=	12
FIRST CONST=	0.00019	XT=	4.0	NT=	12
FIRST CONST=	0.00135	XT=	5.0	NT=	12
FIRST CONST=	0.00736	XT=	6.0	NT=	12
FIRST CONST=	0.03140	XT=	7.0	NT=	12
FIRST CONST=	0.10520	XT=	8.0	NT=	12
SECOND CONST=	0.55275	XT=	7.0	NT=	12
FIRST CONST=	0.00000	XT=	0.0	NT=	13
FIRST CONST=	0.00000	XT=	1.0	NT=	13
FIRST CONST=	0.00000	XT=	2.0	NT=	13
FIRST CONST=	0.00001	XT=	3.0	NT=	13
FIRST CONST=	0.00005	XT=	4.0	NT=	13
FIRST CONST=	0.00041	XT=	5.0	NT=	13
FIRST CONST=	0.00245	XT=	6.0	NT=	13
FIRST CONST=	0.01157	XT=	7.0	NT=	13
FIRST CONST=	0.04379	XT=	8.0	NT=	13
FIRST CONST=	0.132[illegible]	XT=	9.0	NT=	13
SECOND CONST=	0.655[illegible]	XT=	8.0	NT=	13
FIRST CONST=	0.00000	XT=	0.0	NT=	14
FIRST CONST=	0.00000	XT=	1.0	NT=	14
FIRST CONST=	0.00000	XT=	2.0	NT=	14
FIRST CONST=	0.00000	XT=	3.0	NT=	14
FIRST CONST=	0.00002	XT=	4.0	NT=	14
FIRST CONST=	0.00012	XT=	5.0	NT=	14
FIRST CONST=	0.00086	XT=	6.0	NT=	14
FIRST CONST=	0.00419	XT=	7.0	NT=	14
FIRST CONST=	0.01717	XT=	8.0	NT=	14
FIRST CONST=	0.05857	XT=	9.0	NT=	14
FIRST CONST=	0.162[illegible]5	XT=	10.0	NT=	14
SECOND CONST=	0.71435	XT=	9.0	NT=	14
FIRST CONST=	0.00000	XT=	0.0	NT=	15
FIRST CONST=	0.00000	XT=	1.0	NT=	15
FIRST CONST=	0.00000	XT=	2.0	NT=	15
FIRST CONST=	0.00000	XT=	3.0	NT=	15
FIRST CONST=	0.00000	XT=	4.0	NT=	15
FIRST CONST=	0.00004	XT=	5.0	NT=	15
FIRST CONST=	0.00025	XT=	6.0	NT=	15
FIRST CONST=	0.00141	XT=	7.0	NT=	15
FIRST CONST=	0.00645	XT=	8.0	NT=	15
FIRST CONST=	0.02432	XT=	9.0	NT=	15
FIRST CONST=	0.0757[illegible]	XT=	10.0	NT=	15
FIRST CONST=	0.19347	XT=	11.0	NT=	15
SECOND CONST=	0.78264	XT=	10.0	NT=	15
SECOND CONST=	0.59272	XT=	9.0	NT=	15

X= 10.00000 N= 15

APPENDIX B

A Listing of the Program for a Bayesian

Binomial Hypothesis Testing Procedure

```
LIM=1,5
J=TEST
// EXEC FORX2
//FORT.SYSIN DD *
      IMPLICIT REAL*8(A-H,O-Z)
      INTEGER IER
      READ(5,10) DEL,SGM,SDEL,ALF,BETA,NT
   10 FORMAT(5F10.5,I4)
      BET=1.0-BETA
      X1=DEL
      X2=1.0
C WE START THE ALGORITHM BY INITIATING XT AS ZERO
      W1=SGM
      W2=SDEL
      A1=W1
      B1=W2
      CALL FACT1(A1,B1,SON)
      W=SON
   11 XT=0.0
      WNT=NT
      W4=WNT+SDEL
      TA1=SGM
      TB1=W4
      CALL FACT2(TA1,TB1,TERS)
      PAR=TERS
      CO1=W*PAR
C THIS IS THE VALUE WHEN XT IS ZERO
C NOW WE COMPUTE THE VALUE G1 WHEN XT IS OTHER THAN ZERO.
  301 IXT=XT
      TOT=CO1
      IF(XT.EQ.0.0) GO TO 1001
      DO 1000 I=1,IXT
      RI=I
      P1=W1+RI
      P2=W4-RI
      TA1=P1
      TB1=P2
      CALL FACT2(TA1,TB1,TERS)
      P3=WNT+1.0
      P4=P3-RI
      P5=RI+1.0
      Z=(DGAMMA(P3))/((DGAMMA(P4))*(DGAMMA(P5)))
      P=TERS
      TOT=TOT+(P*Z*W)
 1000 CONTINUE
 1001 G1=TOT
      WRITE(6,60) G1,XT,NT
   60 FORMAT(5X,'FIRST CONST=',F10.5,5X,'XT=',F5.1,5X,'NT=',I4)
C SO WE COMPUTED THE VALUE OF FIRST CONSTRAINT
      IF(G1.GT.ALF) GO TO 333
      IF(XT.EQ.NT) GO TO 380
      XT=XT+1.0
      GO TO 301
  333 XT=XT-1.0
```

```
      IF(XT.LT.0.0) GO TO 999
C OTHERWISE WE GO AND CALCULATE G2
  380 WW=W*(DEL**WNT)
C NOW COMPUTE THE VALUE WHEN XT IS ZERO,THAT IS J IS ZERO.
C WHEN J IS ZERO L IS ZERO
C WHEN J IS ZERO,M GOES FROM ZERO TO NT AND L IS ALWAYS ZERO IN THIS CASE
C FIRST CONSIDER THE CASE WHERE WHEN M IS ZERO
      A=W1
      B=W2
      TA1=W1
      TB1=W2
      CALL FACT2(TA1,TB1,TERS)
      CALL MDBETA(X1,A,B,P1,IER)
      CALL MDBETA(X2,A,B,P2,IER)
      Y=TERS
      VALO=(P2-P1)*Y
      SUM=VALO
C NOW CONSIDER THE CASES WHERE M IS ONE TO NT.
      DO 1500 M=1,NT
      A=W1
      BM=M
      BM1=WNT+1.0
      BM2=WNT-BM+1.0
      BM3=BM+1.0
      BMCOM=DGAMMA(BM1)/((DGAMMA(BM2))*(DGAMMA(BM3)))
      BFAC=(DEL**(-BM))*BMCOM
      B=W2+BM
      TA1=W1
      TB1=B
      CALL FACT2(TA1,TB1,TERS)
      CALL MDBETA(X1,A,B,P1,IER)
      CALL MDBETA(X2,A,B,P2,IER)
      Y=TERS
      VAL=(P2-P1)*Y*BFAC
      SUM=SUM+VAL
 1500 CONTINUE
      JXT=XT
      RJSUM=SUM
C IF XT IS ZERO WE HAVE ONLY THE ABOVE TERM
      IF(XT.EQ.0.0) GO TO 2001
      DO 2000 J=1,JXT
C THIS IS THE MOST OUTER SUM
      RJ=J
      RJ1=WNT+1.0
      RJ2=WNT-RJ+1.0
      RJ3=RJ+1.0
      COMBJ=(DGAMMA(RJ1))/((DGAMMA(RJ2))*(DGAMMA(RJ3)))
C NOW L IS FROM ZERO TO J.AGAIN CONSIDER THE CASE WHERE L IS ZERO
      LP=(-1)**J
      PL=LP
C NOTE WHEN L IS ZERO M GOES FROM ZERO TO NT-J
      LJL=NT-J
      IF(LJL.EQ.0) GO TO 2101
      DO 2100 M=1,LJL
```

```
      RRM=M
      RRM1=WNT-RJ+1.0
      RRM2=WNT-RJ-RRM+1.0
      RRM3=RRM+1.0
      RCOM=(DGAMMA(RRM1))/((DGAMMA(RRM2))*(DGAMMA(RRM3)))
      FFAC=(DEL**(-RRM))*RCOM
      A=SGM
      B=RRM+SDEL
      TA1=A
      TB1=B
      CALL FACT2(TA1,TB1,TERS)
      CALL MDBETA(X1,A,B,P1,IER)
      CALL MDBETA(X2,A,B,P2,IER)
      Y=TERS
      VALM=(P2-P1)*FFAC*Y
      VALO=VALO+VALM
 2100 CONTINUE
 2101 RLSUM=VALO*PL
C THIS IS THE VALUE WHEN L IS ZERO
C NOW WE WANT TO CONSIDER L FROM 1 TO J.THIS IS THE SECOND SUM
      DO 2500 L=1,J
      RL=L
      RL1=RJ-RL+1.0
      RL2=RL+1.0
      COMBL=(DGAMMA(RJ3))/((DGAMMA(RL1))*(DGAMMA(RL2)))
      LPL=(-1)**(J-L)
      FLP=LPL
      POWER=DEL**(-RL)
      FACL=FLP*COMBL*POWER
C NOW SHOULD CONSIDER M LOOP AGAIN.NOW M IS FROM ZERO TO NT-J FOR GIVEN L
C START WITH MIS ZERO
      A=RL+SGM
      B=SDEL
      CALL MDBETA(X1,A,B,P1,IER)
      CALL MDBETA(X2,A,B,P2,IER)
      TA1=A
      TB1=B
      CALL FACT2(TA1,TB1,TERS)
      Y=TERS
      VAL=(P2-P1)*Y
      RMSUM=VAL
      LL=NT-J
      IF(LL.EQ.0) GO TO 3001
      DO 3000 M=1,LL
      RM=M
      RM1=WNT-RJ+1.0
      RM2=WNT-RJ-RM+1.0
      RM3=RM+1.0
      COMBM=(DGAMMA(RM1))/((DGAMMA(RM2))*(DGAMMA(RM3)))
      FACM=(DEL**(-RM))*(COMBM)
      A=RL+SGM
      B=RM+SDEL
      CALL MDBETA(X1,A,B,P1,IER)
      CALL MDBETA(X2,A,B,P2,IER)
```

```
      TA1=A
      TB1=B
      CALL FACT2(TA1,TB1,TERS)
      Y=TERS
      VAL=(P2-P1)*FACM*Y
      RMSUM=RMSUM+VAL
 3000 CONTINUE
 3001 RRSUM=RMSUM
C THE MOST INNER LOOP IS FINISHED.
      RLSUM=(FACL*RRSUM)+RLSUM
C THIS IS THE SUM FOR L LOOP
 2500 CONTINUE
C L LOOP IS FINISHED
C NOW FINISH J LOOP.THE MOST OUTER LOOP.
      RJSUM=(COMBJ*RLSUM)+RJSUM
 2000 CONTINUE
C SO WE EVALUATED G2.
 2001 G2=RJSUM*WW
      WRITE(6,61) G2,XT,NT
   61 FORMAT(5X,'SECOND CONST=',F10.5,5X,'XT=',F5.1,5X,'NT=',I4)
      IF(G2.LT.BET) GO TO 999
  777 IF(XT.LT.1.0) GO TO 888
      XT=XT-1.0
C CHECK G2 AGAIN.
      WW=W*(DEL**WNT)
C NOW COMPUTE THE VALUE WHEN XT IS ZERO,THAT IS J IS ZERO.
C WHEN J IS ZERO L IS ZERO.
C WHEN J IS ZERO,M GOES FROM ZERO TO NT AND L IS ALWAYS ZERO IN THIS CASE
C FIRST CONSIDER THE CASE WHERE WHEN M IS ZERO
      A=W1
      B=W2
      CALL MDBETA(X1,A,B,P1,IER)
      CALL MDBETA(X2,A,B,P2,IER)
      TA1=A
      TB1=B
      CALL FACT2(TA1,TB1,TERS)
      Y=TERS
      VALO=(P2-P1)*Y
      SUM=VALO
C NOW CONSIDER THE CASES WHERE M IS ONE TO NT.
      DO 1501 M=1,NT
      A=W1
      BM=M
      BM1=WNT+1.0
      BM2=WNT-BM+1.0
      BM3=BM+1.0
      BMCOM=DGAMMA(BM1)/((DGAMMA(BM2))*(DGAMMA(BM3)))
      BFAC=(DEL**(-BM))*BMCOM
      B=W2+BM
      TA1=A
      TB1=B
      CALL FACT2(TA1,TB1,TERS)
      CALL MDBETA(X1,A,B,P1,IER)
      CALL MDBETA(X2,A,B,P2,IER)
```

```
      Y=TERS
      VAL=(P2-P1)*Y*BFAC
      SUM=SUM+VAL
 1501 CONTINUE
      JXT=XT
      RJSUM=SUM
C IF XT IS ZERO WE HAVE ONLY THE ABOVE TERM
      IF(XT.EQ.0.0) GO TO 2011
      DO 5000 J=1,JXT
C THIS IS THE MOST OUTER SUM
      RJ=J
      RJ1=WNT+1.0
      RJ2=WNT-RJ+1.0
      RJ3=RJ+1.0
      COMBJ=(DGAMMA(RJ1))/((DGAMMA(RJ2))*(DGAMMA(RJ3)))
C NOW L IS FROM ZERO TO J.AGAIN CONSIDER THE CASE WHERE L IS ZERO
      LP=(-1)**J
      PL=LP
C NOTE WHEN L IS ZERO M GOES FROM ZERO TO NT-J
      LJL=NT-J
      IF(LJL.EQ.0) GO TO 2102
      DO 2105 M=1,LJL
      RRM=M
      RRM1=WNT-RJ+1.0
      RRM2=WNT-RJ-RRM+1.0
      RRM3=RRM+1.0
      RCOM=(DGAMMA(RRM1))/((DGAMMA(RRM2))*(DGAMMA(RRM3)))
      FFAC=(DEL**(-RRM))*RCOM
      A=SGM
      B=RRM+SDEL
      CALL MDBETA(X1,A,B,P1,IER)
      CALL MDBETA(X2,A,B,P2,IER)
      TA1=A
      TB1=B
      CALL FACT2(TA1,TB1,TERS)
      Y=TERS
      VALM=(P2-P1)*FFAC*Y
      VALO=VALO+VALM
 2105 CONTINUE
 2102 RLSUM=VALO*PL
C THIS IS THE VALUE WHEN L IS ZERO
C NOW WANT TO CONSIDER L FROM 1 TO J. THIS IS THE SECOND SUM
      DO 2501 L=1,J
      RL=L
      RL1=RJ-RL+1.0
      RL2=RL+1.0
      COMBL=(DGAMMA(RJ3))/((DGAMMA(RL1))*(DGAMMA(RL2)))
      LPL=(-1)**(J-L)
      FLP=LPL
      POWER=DEL**(-RL)
      FACL=FLP*COMBL*POWER
C NOW SHOULD CONSIDER M LOOP AGAIN.NOW M IS FROM ZERO TO NT-J FOR GIVEN L
C START WITH MIS ZERO.
      A=RL+SGM
```

```
      B=SDEL
      CALL MDBETA(X1,A,B,P1,IER)
      CALL MDBETA(X2,A,B,P2,IER)
      TA1=A
      TB1=B
      CALL FACT2(TA1,TB1,TERS)
      Y=TERS
      VAL=(P2-P1)*Y
      RMSUM=VAL
      LL=NT-J
      IF(LL.EQ.0) GO TO 4001
      DO 4000 M=1,LL
      RM=M
      RM1=WNT-RJ+1.0
      RM2=WNT-RJ-RM+1.0
      RM3=RM+1.0
      COMBM=(DGAMMA(RM1))/((DGAMMA(RM2))*(DGAMMA(RM3)))
      FACM=(DEL**(-RM))*(COMBM)
      A=RL+SGM
      B=RM+SDEL
      CALL MDBETA(X1,A,B,P1,IER)
      CALL MDBETA(X2,A,B,P2,IER)
      TA1=A
      TB1=B
      CALL FACT2(TA1,TB1,TERS)
      Y=TERS
      VAL=(P2-P1)*FACM*Y
      RMSUM=RMSUM+VAL
 4000 CONTINUE
 4001 RRSUM=RMSUM
C THE MOST INNER LOOP IS FINISHED.
      RLSUM=(FACL*RRSUM)+RLSUM
C THIS IS THE SUM FOR L LOOP
 2501 CONTINUE
C L LOOP IS FINISHED.
C NOW FINISH J LOOP. THE MOST OUTER LOOP.
      RJSUM=(COMBJ*RLSUM)+RJSUM
 5000 CONTINUE
C SO WE EVALUATED G2.
 2011 G2=RJSUM*WW
      WRITE(6,62) G2,XT,NT
   62 FORMAT(5X,'SECOND CONST='F10.5,5X,'XT=',F5.1,5X,'NT=',I4)
C CHECK G2 NOW
      IF(G2.GE.BET) GO TO 777
      XT=XT+1.0
      GO TO 888
  999 NT=NT+1
      GO TO 11
  888 WRITE(6,555) XT,NT
  555 FORMAT(10X,'X=',F10.5,5X,'N=',I4)
      STOP
      END
      SUBROUTINE FACT1(A1,B1,SON)
      IMPLICIT REAL*8(A-H,O-Z)
```

```
      C=A1+B1
      IF(A1.LE.57.0.AND.C.LE.57.0) GO TO 41
      C1=C-1.0
      A2=A1-1.0
      B2=B1-1.0
      C2=A2+B2
      IB=A2+1.0
      IC=C2
      PAY=C1
      DO 42 I=IB,IC
      ZI=I
      PAY=PAY*ZI
   42 CONTINUE
      PAYDA=1.0
      JA=B2
      DO 43 J=1,JA
      VJ=J
      PAYDA=PAYDA*VJ
   43 CONTINUE
      SON=PAY/PAYDA
      GO TO 45
   41 SON=DGAMMA(C)/((DGAMMA(A1))*(DGAMMA(B1)))
   45 CONTINUE
      RETURN
      END
      SUBROUTINE FACT2(TA1,TB1,TERS)
      IMPLICIT REAL*8(A-H,O-Z)
      C=TA1+TB1
      IF(TA1.LE.57.0.AND.C.LE.57.0) GO TO 71
      C1=C-1.0
      A2=TA1-1.0
      B2=TB1-1.0
      C2=A2+B2
      IB=A2+1.0
      IC=C2
      PAY=C1
      DO 72 I=IB,IC
      ZI=I
      PAY=PAY*ZI
   72 CONTINUE
      PAYDA=1.0
      JA=B2
      DO 73 J=1,JA
      VJ=J
      PAYDA=PAYDA*VJ
   73 CONTINUE
      TERS=PAYDA/PAY
      GO TO 75
   71 TERS=((DGAMMA(TA1))*(DGAMMA(TB1)))/(DGAMMA(C))
   75 CONTINUE
      RETURN
      END
//GO.SYSLIB DD
//     DD     DSN=GWU.IMSL.V9.DLOAD,DISP=SHR
```

```
//GO.SYSIN    DD      *
   0.25000 113.00000  20.00000   0.10100   0.25000   1
//
```

APPENDIX C

Illustrative Calculation of Expected Sample Sizes for Curtailed Sequential Sampling

C.1. THE CASE OF TESTING ONE ITEM AT A TIME

We illustrate this for Stage 0. Here $x_j^* = 5$, $n_t = 17$.

We must have either 6 successes to accept, or 12 failures to reject

$$P[n_t=6|p_t] = \binom{5}{5} p_t^6 = 0.015625$$

$$P[n_t=7|p_t] = \binom{6}{5} p_t^6(1-p_t) = 0.046875$$

$$P[n_t=8|p_t] = \binom{7}{5} p_t^6(1-p_t)^2 = 0.0820312$$

$$P[n_t=9|p_t] = \binom{8}{5} p_t^6(1-p_t)^3 = 0.109375$$

$$P[n_t=10|p_t] = \binom{9}{5} p_t^6(1-p_t)^4 = 0.1230469$$

$$P[n_t=11|p_t] = \binom{10}{5} p_t^6(1-p_t)^5 = 0.1230469$$

$$P[n_t=12|p_t] = \binom{11}{5} p_t^6(1-p_t)^6 + \binom{11}{11} (1-p_t)^{12} = 0.1130371$$

$$P[n_t=13|p_t] = \binom{12}{5} p_t^6(1-p_t)^7 + \binom{12}{11} p_t(1-p_t)^{12} = 0.0968018$$

$$P[n_t=14|p_t] = \binom{13}{5} p_t^6(1-p_t)^8 + \binom{13}{11} p_t^2(1-p_t)^{12} = 0.083313$$

$$P[n_t=15|p_t] = \binom{14}{5} p_t^6(1-p_t)^9 + \binom{14}{11} p_t^3(1-p_t)^{12} = 0.0722046$$

$$P[n_t=16|p_t] = \binom{15}{5} p_t^6(1-p_t)^{10} + \binom{15}{11} p_t^4(1-p_t)^{12} = 0.0666504$$

$$P[n_t=17|p_t] = \binom{16}{5} p_t^6(1-p_t)^{11} + \binom{16}{11} p_t^5(1-p_t)^{12} = 0.0666504$$

To obtain $P[n_t=j]$, $j = 6,7,\dots,17$, we average out the above by using $g(p_t|\cdot)$. At Stage 0, $\gamma = 1$, $\delta = 1$.

$$p[n_t=6] = \frac{\Gamma(\gamma+6)\Gamma(\delta)}{\Gamma(\gamma+\delta+6)} = 0.1428571$$

$$p[n_t=7] = 6\,\frac{\Gamma(\gamma+6)\Gamma(\delta+1)}{\Gamma(\gamma+\delta+7)} = 0.1071429$$

$$p[n_t=8] = 21\,\frac{\Gamma(\gamma+6)\Gamma(\delta+2)}{\Gamma(\gamma+\delta+8)} = 0.083333$$

$$p[n_t=9] = 56\,\frac{\Gamma(\gamma+6)\Gamma(\delta+3)}{\Gamma(\gamma+\delta+9)} = 0.0666667$$

$$p[n_t=10] = 126\,\frac{\Gamma(\gamma+6)\Gamma(\delta+4)}{\Gamma(\gamma+\delta+10)} = 0.0545455$$

$$p[n_t=11] = 252\,\frac{\Gamma(\gamma+6)\Gamma(\delta+5)}{\Gamma(\gamma+\delta+11)} = 0.0454545$$

$$p[n_t=12] = 402\,\frac{\Gamma(\gamma+6)\Gamma(\delta+6)}{\Gamma(\gamma+\delta+12)} + \frac{\Gamma(\gamma)\Gamma(\delta+12)}{\Gamma(\gamma+\delta+12)} = 0.1103896$$

$$p[n_t=13] = 792\,\frac{\Gamma(\gamma+6)\Gamma(\delta+7)}{\Gamma(\gamma+\delta+13)} + 12\,\frac{\Gamma(\gamma+1)\Gamma(\delta+12)}{\Gamma(\gamma+\delta+13)} = 0.0989011$$

$$p[n_t=14] = 1287\,\frac{\Gamma(\gamma+6)\Gamma(\delta+8)}{\Gamma(\gamma+\delta+14)} + 78\,\frac{\Gamma(\gamma+2)\Gamma(\delta+12)}{\Gamma(\gamma+\delta+14)} = 0.0857143$$

$$p[n_t=15] = 2002\,\frac{\Gamma(\gamma+6)\Gamma(\delta+9)}{\Gamma(\gamma+\delta+15)} + 364\,\frac{\Gamma(\gamma+3)\Gamma(\delta+12)}{\Gamma(\gamma+\delta+15)} = 0.075$$

$$p[n_t=16] = 3003\,\frac{\Gamma(\gamma+6)\Gamma(\delta+10)}{\Gamma(\gamma+\delta+16)} + 1365\,\frac{\Gamma(\gamma+4)\Gamma(\delta+12)}{\Gamma(\gamma+\delta+16)} = 0.066176$$

$$p[n_t=17] = 4368\,\frac{\Gamma(\gamma+6)\Gamma(\delta+11)}{\Gamma(\gamma+\delta+17)} + 4368\,\frac{\Gamma(\gamma+5)\Gamma(\delta+12)}{\Gamma(\gamma+\delta+17)} = 0.0588235$$

$$E[n_t] = 10.91\ .$$

C.2. THE CASE OF TESTING IN BATCHES OF SIZE 3

Stage 0

$$p_t = 0.5 \quad (1-p_t) = 0.5$$

$$x_j^* = 5 \qquad n_t = 17$$

We must have either 6 successes to accept or 12 failures to reject.

Thus, $n_t \varepsilon \{6,9,12,15.18\}$

$$p[n_t = 6|p_t] = \binom{6}{6} p_t^6 = 0.015625$$

$$p[n_t=9|p_t] = \binom{6}{5} p_t^6(1-p_t) + \binom{7}{5} p_t^6(1-p_t)^2 + \binom{8}{5} p_t^6(1-p_t)^3 = 0.2382812$$

$$p[n_t=12|p_t] = \binom{9}{5} p_t^6(1-p_t)^4 + \binom{10}{5} p_t^6(1-p_t)^5 + \binom{11}{5} p_t^6(1-p_t)^6$$

$$+ \binom{12}{12} (1-p_t)^{12} = 0.3444824$$

$$p[n_t=15|p_t] = \binom{12}{5} p_t^6(1-p_t)^7 + \binom{13}{5} p_t^6(1-p_t)^8 + \binom{14}{5} p_t^6(1-p_t)^9$$

$$+ \binom{12}{11} p_t(1-p_t)^{12} + \binom{13}{11} p_t^2(1-p_t)^{12} + \binom{14}{11} p_t^3(1-p_t)^{12} = 0.2536621$$

$$p[n_t=18|p_t] = \binom{15}{15} p_t^5(1-p_t)^{10} + \binom{15}{4} p_t^4(1-p_t)^{11} = 0.1333008$$

Stage 1

$$p_t = 0.875 \quad (1-p_t) = 0.125$$

$$x_t^* = 9 \qquad n_t = 13$$

We must have either 10 successes to accept or 4 failures to reject.

Thus, $n_t \varepsilon \{6,9,12,15\}$

$$p[n_t=6|p_t] = \binom{6}{4} p_t^2(1-p_t)^4 + \binom{6}{5} p_t(1-p_t)^5 + \binom{6}{6} (1-p_t)^6 = 0.0029678$$

$$p[n_t=9|p_t] = \binom{6}{3} p_t^3(1-p_t)^4 + \binom{7}{3} p_t^4(1-p_t)^4 + \binom{8}{3} p_t^5(1-p_t)^4 = 0.0140249$$

$$p[n_t=12|p_t] = \binom{9}{3} p_t^6(1-p_t)^4 + \binom{10}{3} p_t^7(1-p_t)^4 + \binom{11}{3} p_t^8(1-p_t)^4 + \binom{9}{9} p_t^{10}$$
$$+ \binom{10}{9} p_t^{10}(1-p_t) + \binom{11}{9} p_t^{10}(1-p_t)^2 = 0.852551$$

$$p[n_t=15|p_t] = \binom{12}{3} p_t^9(1-p_t)^3 = 0.1291889$$

Stage 2

$$p_t = 0.9 \quad (1-p_t) = 0.1$$

$$x_t^* = 8 \qquad n_t = 11$$

We must have either 9 successes to accept or 3 failures to reject.

Thus, $n_t \varepsilon \{3,6,9,12\}$

$$p[n_t=3|p_t] = \binom{3}{3} (1-p_t)^3 = 0.001$$

$$p[n_t=6|p_t] = \binom{3}{2} p_t(1-p_t)^3 + \binom{4}{2} p_t^2(1-p_t)^3 + \binom{5}{2} p_t^3(1-p_t)^3 = 0.01485$$

$$p[n_t=9|p_t] = \binom{6}{2} p_t^4(1-p_t)^3 + \binom{7}{2} p_t^5(1-p_t)^3 + \binom{8}{2} p_t^6(1-p_t)^3 + \binom{9}{0} p_t^9 = 0.4245426$$

$$p[n_t=12|p_t] = \binom{9}{2} p_t^7(1-p_t)^2 + \binom{9}{1} p_t^8(1-p_t) = 0.5596074$$

Stage 3

$$p_t = 0.906 \quad 1-p_t = 0.094$$

$$x_t^* = 8 \qquad n_t = 11$$

The same enumeration as in Stage 2.

Stage 4

$$p_t = 0.909 \quad 1-p_t = 0.091$$

$$x_t^* = 8 \qquad n_t = 11$$

The same enumeration as in Stage 2.

Stage 5

$$p_t = 0.875 \quad 1 - p_t = 0.125$$

$$x_t^* = 9 \qquad n_t = 13$$

The same enumeration as in Stage 1.

Stage 6

$$p_t = 0.853 \quad 1 - p_t = 0.147$$

$$x_t^* = 8 \qquad n_t = 12$$

We must have either 4 failures to reject or 9 successes to accept.

Thus, $n_t \varepsilon \{6,9,12\}$

$$p[n_t=6|p_t] = \binom{6}{4} p_t^2(1-p_t)^4 + \binom{6}{5} p_t(1-p_t)^5 + \binom{6}{6} (1-p_t)^6 = 0.0054577$$

$$p[n_t=9|p_t] = \binom{6}{3} p_t^3(1-p_t)^4 + \binom{7}{3} p_t^4(1-p_t)^4 + \binom{8}{3} p_t^5(1-p_t)^4 + \binom{9}{0} p_t^9 = 0.2653362$$

$$p[n_t=12|p_t] = \binom{9}{3} p_t^6(1-p_t)^3 + \binom{9}{2} p_t^7(1-p_t)^2 + \binom{9}{1} p_t^8(1-p_t) = 0.7292061$$

Stage 7

$$p_t = 0.825 \quad (1-p_t) = 0.175$$

$$x_j^* = 9 \qquad n_t = 14$$

We must have either 10 successes to accept or 5 failures to reject.

Thus, $n_t \varepsilon \{6,9,12,15\}$

$$p[n_t=6|p_t] = \binom{6}{5} p_t(1-p_t)^5 + \binom{6}{6} (1-p_t)^6 = 0.0008412$$

$$p[n_t=9|p_t] = \binom{6}{4} p_t^2(1-p_t)^5 + \binom{7}{4} p_t^3(1-p_t)^5 + \binom{8}{4} p_t^4(1-p_t)^5 = 0.0102237$$

$$p[n_t=12|p_t] = \binom{9}{4} p_t^5(1-p_t)^5 + \binom{10}{4} p_t^6(1-p_t)^5 + \binom{11}{4} p_t^7(1-p_t)^5 + \binom{9}{9} p_t^{10}$$

$$+ \binom{10}{9} p_t^{10}(1-p_t) + \binom{11}{9} p_t^{10}(1-p_t)^2 = 0.6805573$$

$$p[n_t=15|p_t] = \binom{12}{4} p_t^8(1-p_t)^4 + \binom{12}{9} p_t^9(1-p_t)^3 = 0.3083778$$

<u>Stage 8</u>

$$p_t = 0.833 \quad (1-p_t) = 0.167$$

$$x_t^* = 9 \qquad n_t = 14$$

The same enumeration as in Stage 7.

<u>Stage 9</u>

$$p_t = 0.820 \quad (1-p_t) = 0.180$$

$$x_t^* = 9 \qquad n_t = 14$$

The same enumeration as in Stage 7.

<u>Stage 10</u>

$$p_t = 0.837 \quad (1-p_t) = 0.163$$

$$x_t^* = 9 \qquad n_t = 14$$

The same enumeration as in Stage 7.

<u>Stage 11</u>

$$p_t = 0.841 \quad (1-p_t) = 0.159$$

$$x_t^* = 10 \qquad n_t = 15$$

We must have either 11 successes to accept or 5 failures to reject.

Thus, $n_t \varepsilon \{6,9,12,15\}$

$$p[n_t=6|p_t] = \binom{6}{5} p_t(1-p_t)^5 + \binom{6}{6} (1-p_t)^6 = 0.0005289$$

$$p[n_t=9|p_t] = \binom{6}{4} p_t^2(1-p_t)^5 + \binom{7}{4} p_t^3(1-p_t)^5 + \binom{8}{4} p_t^4(1-p_t)^5 = 0.0067523$$

$$p[n_t=12|p_t] = \binom{9}{4} p_t^5(1-p_t)^5 + \binom{10}{4} p_t^6(1-p_t)^5 + \binom{11}{4} p_t^7(1-p_t)^5 + \binom{10}{10} p_t^{11}$$
$$+ \binom{11}{10} p_t^{11}(1-p_t) = 0.4321114$$

$$p[n_t=15|p_t] = \binom{12}{4} p_t^8(1-p_t)^4 + \binom{12}{9} p_t^9(1-p_t)^3 + \binom{12}{10} p_t^{10} (1-p_t)^2 = 0.5606073$$

<u>Stage 12</u>

$$p_t = 0.836 \quad (1-p_t) = 0.164$$

$$x_t^* = 9 \qquad n_t = 14$$

The same enumeration as in Stage 7.

<u>Stage 13</u>

$$p_t = 0.848 \quad (1-p_t) = 0.152$$

$$x_t^* = 8 \qquad n_t = 12$$

The same enumeration as in Stage 6.

<u>Stage 14</u>

$$p_t = 0.850 \quad (1-p_t) = 0.150$$

$$x_t^* = 8 \qquad n_t = 12$$

The same enumeration as in Stage 6.

To obtain the $E(n_t)$, we average out the above by using $g(p_t|\cdot)$.

We illustrate this for Stage 0.

$$p[n_t=6] = \frac{\Gamma(\gamma+\delta)}{\Gamma(\gamma)\Gamma(\delta)} \frac{\Gamma(\gamma+6)\Gamma(\delta)}{\Gamma(\gamma+\delta+6)} = 0.1428571$$

$$p[n_t=9] = \frac{\Gamma(\gamma+\delta)}{\Gamma(\gamma)\Gamma(\delta)} \left[6 \frac{\Gamma(\gamma+6)\Gamma(\delta+1)}{\Gamma(\gamma+\delta+7)} + 21 \frac{\Gamma(\gamma+6)\Gamma(\delta+2)}{\Gamma(\gamma+\delta+8)} + 56 \frac{\Gamma(\gamma+6)\Gamma(\delta+3)}{\Gamma(\gamma+\delta+0)}\right] = 0.2571429$$

$$p[n_t=12] = \frac{\Gamma(\gamma+\delta)}{\Gamma(\gamma)\Gamma(\delta)} \left[126 \frac{\Gamma(\gamma+6)\Gamma(\delta+4)}{\Gamma(\gamma+\delta+10)} + 252 \frac{\Gamma(\gamma+6)\Gamma(\delta+5)}{\Gamma(\gamma+\delta+11)} + 402 \frac{\Gamma(\gamma+6)\Gamma(\delta+6)}{\Gamma(\gamma+\delta+12)} + \frac{\Gamma(\gamma)\Gamma(\delta+12)}{\Gamma(\gamma+\delta+12)}\right] = 0.2103897$$

$$p[n_t=15] = \frac{\Gamma(\gamma+\delta)}{\Gamma(\gamma)\Gamma(\delta)} \left[792 \frac{\Gamma(\gamma+6)\Gamma(\gamma+7)}{\Gamma(\gamma+\delta+13)} + 1287 \frac{\Gamma(\gamma+6)\Gamma(\delta+8)}{\Gamma(\gamma+\delta+14)} + 2002 \frac{\Gamma(\gamma+6)\Gamma(\delta+9)}{\Gamma(\gamma+\delta+15)} + 12 \frac{\Gamma(\gamma+1)\Gamma(\delta+12)}{\Gamma(\gamma+\delta+13)} + 78 \frac{\Gamma(\gamma+2)\Gamma(\delta+12)}{\Gamma(\gamma+\delta+14)} + 364 \frac{\Gamma(\gamma+3)\Gamma(\delta+12)}{\Gamma(\gamma+\delta+15)}\right] = 0.2596154$$

$$p[n_t=18] = \frac{\Gamma(\gamma+\delta)}{\Gamma(\gamma)\Gamma(\delta)} \left[3003 \frac{\Gamma(\gamma+5)\Gamma(\delta+10)}{\Gamma(\gamma+\delta+15)} + 1365 \frac{\Gamma(\gamma+4)\Gamma(\delta+11)}{\Gamma(\gamma+\delta+15)}\right] = 0.125$$

$E[n_t] = 11.84$.

Similarly, we can obtain $E[n_t]$ for other stages.

REFERENCES

BOX, G. E. P. (1982). An apology ecumenism in statistics, MRC Technical Report #2408.

ACKNOWLEDGMENTS

The algorithm of Appendix A is due to Mr. Jerzy Kyparisis, and the computer code was developed by Mr. Refik Soyer. We have benefited from discussions with Professors Box, Draper, Harris, Johnson, Leonard and Woodroofe. The problem was brought to our attention by Dr. Willard of the Office of the Deputy Undersecretary of the Army; his comments have helped us develop the methodology proposed. Comments by Dr. Moore of BRL are also appreciated.